

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
25 September 2003 (25.09.2003)

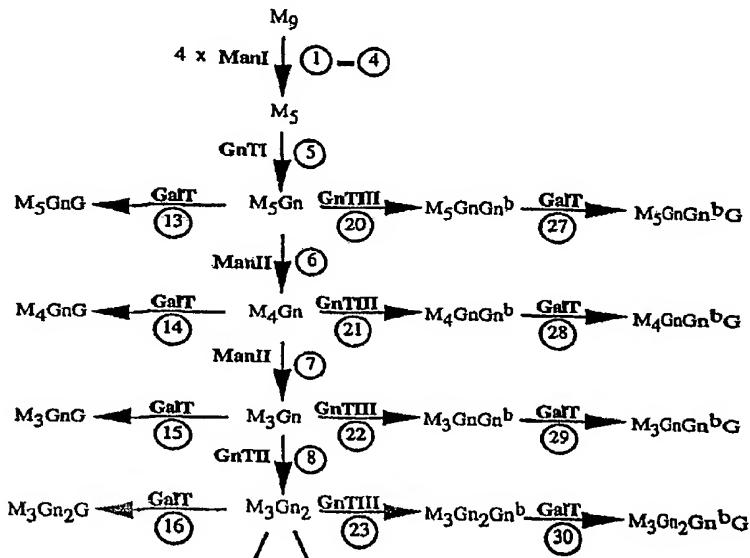
PCT

(10) International Publication Number
WO 03/078614 A2

- (51) International Patent Classification⁷: **C12N 9/00**
- (21) International Application Number: **PCT/IB03/01562**
- (22) International Filing Date: 18 March 2003 (18.03.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/365,769 19 March 2002 (19.03.2002) US
60/368,047 26 March 2002 (26.03.2002) US
- (71) Applicant (for all designated States except US): **PLANT RESEARCH INTERNATIONAL B.V. [NL/NL]**
Droevedaalsesteeg 1, NL-6708 PB Wageningen (NL).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **BAKKER, Hendrikus, Antonius, Cornelius [DE/DE]**; Drachenfeld 75, NL30627 Hanover (DE). **FLORACK, Dionisius, Elisabeth, Antonius [NL/NL]**; Van Doesburglaan 160, NL-6708 MD Wageningen (NL). **BOSCH, Hendrik,**
- Jan [NL/NL]; Win Sonneveldstraat 17, NL-6708 NA Wageningen (NL).
- (74) Agents: **MARSHALL, Cameron, J. et al.**; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: GNTIII EXPRESSION IN PLANTS

**WO 03/078614 A2**

(57) **Abstract:** The invention relates to the field of glycoprotein processing in transgenic plants used as cost efficient and contamination safe factories for the production of recombinant biopharmaceutical proteins or pharmaceutical compositions comprising these glycoproteins. The invention provides a plant comprising a functional mammalian enzyme providing mammalian GnTIII that is normally not present in plants, said plant additionally comprising at least a second mammalian protein or functional fragment thereof that is normally not present in plants.



Published:

- without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

GNTH EXPRESSION IN PLANTS

FIELD OF THE INVENTION

The invention relates to expression of a mammalian N-acetylglucosaminyl-transferase III (GnTIII) enzyme in plants and its use in producing glycoproteins with bisected oligosaccharides and increased amount of terminal G1cNAc residues. The invention further relates to a hybrid protein comprising the catalytic site of GnTIII and transmembrane domain of Golgi apparatus and/or endoplasmic reticulum (ER) protein or modified GNTH comprising ER retention signals and its use in producing glycoproteins with oligosaccharides that lack immunogenic xylose and fucose residues.

10

BACKGROUND OF THE INVENTION

N-Acetylglucosaminyltransferases (GlcNAc-transferases) are "branching" enzymes that add an N-acetylglucosamine (G1cNAc) residue to one of the mannoses of the trimannosyl core structure of typical N-linked glycans. At least six G1cNAc-transferases are known with little or no sequence homology. Besides different protein structures, these G1cNAc-transferases also have different enzymatic properties and substrate specificity. All are typical type II transmembrane proteins with a cytoplasmic domain, a transmembrane anchor and an extracellular stem region with catalytic domain.

A remarkable G1cNAc-transferase is G1cNAc-transferase III (GnTIII). GnTIII, also known as UDP-N-acetylglucosamine:β-D-mannoside β(1,4)-N-acetylglucosaminyl-transferase III (EC 2.4.1.144), inserts bisecting G1cNAc residues in complex-type N-linked glycans of cellular glycoproteins (for a review see Taniguchi, *et al.*, "A glycomic approach to the identification and characterization of glycoprotein function in cells transfected with glycosyltransferase genes" *Proteomics* 1:239247, 2001). GnTIII adds the G1cNAc through a β(1,4) linkage to the β-linked mannose of the trimannosyl core structure of the N-linked glycan. GnTIII was first identified in hen oviduct (Narasimhan S., "Control of glycoprotein synthesis. UDP-G1cNAc:glycopeptide β 4-N-acetylglucosaminyltransferase III, an enzyme in hen oviduct which adds G1cNAc in β14 linkage to the β-linked mannose of the trimannosyl core of N-glycosyl oligosaccharides" *The Journal of Biological Chemistry* 257:10235-10242, 1982) but a high level of activity has also been reported in various types of rat hepatomas, human serum, liver and hepatoma tissues of patients with hepatomas and liver cirrhosis (Ishibashi, *et al.*, "N-acetylglucosaminyltransferase III in human serum and liver and hepatoma tissues: increased activity in liver cirrhosis and hepatoma patients" *Clinical Chimica Acta* 185:325, 1989; Narishimhan, *et al.*, "Expression of N-acetylglucosaminyltransferase III in hepatic nodules during rat liver carcinogenesis promoted by orotic acid" *Journal of Biological Chemistry* 263:1273-1281, 1988; Nishikawa, *et al.* "Determination of N-acetylglucosaminyltransferases III, IV and V in normal and hepatoma tissues of rats" *Biochimica et Biophysica Acta* 1035:313-318, 1990; Pascale, *et al.*, "Expression of N-acetylglucosaminyltransferase III in hepatic nodules generated by different models of rat liver carcinogenesis" *Carcinogenesis* 10:961964, 1989). Bisected oligosaccharides on glycoproteins have

been implicated in antibody-dependent cellular cytotoxicity (ADCC). ADCC is a lytic attack on antibody-targeted cells and is triggered upon binding of lymphocyte receptors to the constant region (Fc) of antibodies. Controlled expression of GnTIII in recombinant Chinese Hamster Ovary (CHO) production cell lines that lack GnTIII activity resulted in antibodies with bisected oligosaccharides with optimized ADCC activity (Davies, *et al.*, "Expression of GnTIII in a recombinant anti-CD20 CHO production cell line: expression of antibodies with altered glycoforms leads to an increase in ADCC through higher affinity for Fc γ RIII" *Biotechnology and Bioengineering* 74:288-294, 2001; Umana, *et al.*, "Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity" *Nature Biotechnology* 17:176-180, 1999). The ADCC activity correlated well with the level of Fc region-associated bisected complex oligosaccharides present on the recombinant antibody (Umana, *et al.*, "Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity" *Nature Biotechnology* 17:176-180, 1999). Bisecting GlcNAc residues resulting from GnTIII activity affect the conformation of the sugar chains in such a way that other glycosyltransferases such as GlcNAc-transferase II and α 1,6-fucosyltransferase, but not β (1,4)-galactosyltransferase, can no longer act (Taniguchi, *et al.*, 2001). Overexpression of GnTIII in CHO cells is lethal.

In contrast to typical mammalian production cell lines such as CHO cells, transgenic plants are generally recognized as a safe production system for therapeutic proteins. Plant glycoproteins, however, differ in oligosaccharide structure with those from mammals in several aspects. They lack terminal galactose and sialic acid, have an additional core xylose and differently linked core fucose (α -1,3) instead of (α -1,6). Like CHO and other pharmaceutical production cell lines they also completely lack bisected oligosaccharides. Plants have the capacity to generate the common core structure, GN2M3GN2 but predominantly M3 GN2 variants are found, indicating removal of terminal GN by hexosaminidases.

Biogenesis of N-linked glycans begins with the synthesis of a lipid linked oligosaccharide moiety (Glc3Man9GlcNAc2-) which is transferred *en bloc* to the nascent polypeptide chain in the endoplasmic reticulum (ER). Through a series of trimming reactions by exoglycosidases in the ER and cis-Golgi compartments the so-called "high mannose" (Man9GlcNAc2 to Man5GlcNAc2) glycans are formed. Subsequently, the formation of complex type glycans starts with the transfer of the first GlcNAc onto Man5GlcNAc2 by GnTI and further trimming by mannosidase II (Mann) to form GlcNAcMan3 GlcNAc2. Complex glycan biosynthesis continues while the glycoprotein is progressing through the secretory pathway with the transfer in the Golgi apparatus of the second GlcNAc residue by GnTII as well as other monosaccharide residues onto the GlcNAcMan3GlcNAc2 under the action of several other glycosyl transferases. Plants and mammals differ with respect to the formation of complex glycans. In plants, complex glycans are characterized by the presence of β (1,2)-xylose residues linked to the Man-3 and/or an α (1,3)-fucose residue linked to GlcNAc1, instead of an α (1,6)-fucose residue linked to the GlcNAc-1 (Lerouge, P., *et al.*, "N-glycoprotein biosynthesis in plants: recent developments and future trends" *Plant Mol Biol* 38:31-48, 1998). Genes encoding the corresponding xylosyl (XylT) and fucosyl (FucT) transferases have been

isolated (Strasser R, "Molecular cloning and functional expression of β 1, 2-xylosyltransferase cDNA from *Arabidopsis thaliana*" *FEBS Lett.* 472:105-8, 2000; Leiter, H., *et al.*, "Purification, cDNA cloning, and expression of GDP-L-Fuc:Asn-linked G1cNAc α 1,3-fucosyltransferase from mung beans" *J Biol Chem.* 274:21830, 1999). Xylose and fucose epitopes are known to be highly immunogenic and possibly allergenic which may pose a problem when plant are used for the production of therapeutic glycoproteins. Moreover, blood serum of many allergy patients contains IgE directed against these epitopes which make particularly these patients at risk to treatments with xylose and fucose containing recombinant proteins. In addition, this carbohydrate directed IgE in sera might cause false positive reaction in *in vitro* tests using plant extracts since there is evidence that these carbohydrate specific IgE's are not relevant for the allergenic reaction. Plants do not possess β (1,4)galactosyltransferases nor α (2,6)sialyltransferases and consequently plant glycans lack the β (1,4)galactose and terminal α (2,6)NeuAc residues often found on mammalian glycans (Vitale and Chrispeels, "Transient N-acetylglucosamine in the biosynthesis of phytohemagglutinin: attachment in the Golgi apparatus and removal in protein bodies" *J Cell Biol* 99:133-140, 1984; Lerouge, P., *et al.*, "N-glycoprotein biosynthesis in plants: recent developments and future trends" *Plant Mol Biol* 38:31-48, 1998).

The final glycan structures are not only determined by the mere presence of enzymes involved in their biosynthesis but to a large extend by the specific sequence of the various enzymatic reactions. The latter is controlled by discrete sequestering and relative position of these enzymes throughout the ER and Golgi, which is mediated by the interaction of determinants of the transferase and specific characteristics of the sub-Golgi compartment for which the transferase is destined. A number of studies using hybrid, molecules have identified that the transmembrane domains of several glycosyltransferases play a central role in their sub-Golgi sorting (Grabenhorst E., *et.al.*, *J. Biol. Chem.* 274:36107-36116, 1999; Colley, K., *Glycobiology* 7:1-13, 1997, Munro, S., *Trends Cell Biol.* 8:11-15, 1998; Gleeson P.A., *Histochem. Cell Biol.* 109:517-532, 1998).

Similar to mammalian production cell lines used in pharmaceutical industry, glycoproteins produced in plants lack GnTIII activity. Plants not only lack GnTIII activity but are completely devoid of GnTIII-like sequences. In addition, plants also lack GnTIV, GnTV ands GnTVI sequences and moreover, sialic acid residues. (For an overview of the major glycosylation attributes of commonly used cell expression systems including plants see, Jenkins, *et al.*, "Getting the glycosylation right: implications for the biotechnology industry" *Nature Biotechnology* 14:975-979, 1996). Nevertheless, plants are a very potent production system. Plants are generally accepted as safe and are free of particles infectious to humans. Plant production is easy scalable and N-linked glycosylation can be controlled (Bakker, *et al.*, "Galactose-extended glycans of antibodies produced by transgenic plants" *Proc. Nat. Acad. Sci. USA* 98:2899-2904, 2001).

Transgenic tobacco plants that produce galactosylated recombinant monoclonal antibodies (Mabs) upon introduction of the human gene for β (1,4)-galactosyltransferase have been reported (hGa1T; Bakker, *et al.*, "Galactose-extended glycans of antibodies produced by transgenic plants" *Proc. Nat. Acad. Sci. USA* 98:2899-2904, 2001; WO01/31044 and WO01/31045).

Therapeutic glycoproteins can be improved by altering their glycosylation pattern (Davies, *et al.*, "Expression of GnTIII in a recombinant anti-CD20 CHO production cell line: expression of antibodies with altered glycoforms leads to an increase in ADCC through higher affinity for Fc γ RIII" *Biotechnology and Bioengineering* 74:288-294, 2001; Umana, *et al.*, "Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity" *Nature Biotechnology* 17:176-180, 1999; Fukuta, *et al.*, "Remodeling of sugar chain structures of human interferon- γ " *Glycobiology* 10:421-430, 2000; Misaizu, *et al.*, "Role of antennary structure of N-linked sugar chains in renal handling of recombinant human erythropoietin" *Blood* 86:4097-4104, 1995; Sburlati, *et al.*, "Synthesis of bisected glycoforms of recombinant IFN- β by overexpression of β -1,4-N-acetylglucosaminyl-transferase III in Chinese Hamster Ovary cells" *Biotechnology Prog.* 14:189-192, 1998). Higher oligosaccharide antennarity of EPO, for example, leads to increased *in vivo* activity due to reduced kidney filtration (Misaizu, *et al.*, "Role of antennary structure of N-linked sugar chains in renal handling of recombinant human erythropoietin" *Blood* 86:4097-4104, 1995). Biosynthesis of such superior glycoforms can be achieved with the "standard" glycosylation machinery of normal production cell lines by two methodologies. The first is by enriching specific glycoforms during purification and the second is by introducing mutations in the polypeptide chain. The latter makes it possible to shift the glycosylation site within the glycoprotein resulting in different glycosylation patterns as the result of differences in accessibility. A complementary route is through genetic engineering of the production cell line itself. New glycosylation patterns can be obtained through expression of glycosyltransferase and glycosidase genes in production cell lines. These genes code for enzymes that either add or remove specific saccharides to and from the glycan of cellular glycoproteins. Several glycosyltransferase genes have been introduced in CHO cells to manipulate glycoform biosynthesis. One of them is GnTIII. Glycosyltransferase GnTIII is involved in branching of the N-linked glycan and results in bisecting GlcNAc residues. CHO cells and other production cell lines typically lack GnTIII activity (Stanley, P. and C.A. Campbell, "A dominant mutation to ricin resistance in chinese hamster ovary cells induces UDP-GlcNAc: glycopeptide β -4-N-acetylglucosaminyl-transferase III activity" *Journal of Biological Chemistry* 261:13370-13378, 1984). Expression of GnTIII in CHO resulted in bisected complex oligosaccharides as expected but overexpression resulted in growth inhibition and was toxic to cells. Similarly, overexpression of GnTV, another glycosyltransferase that introduces triantennary sugar chains, also resulted in growth inhibition suggesting that this may be a general feature of glycosyltransferase overexpression (Umana, *et al.*, "Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity" *Nature Biotechnology* 17:176-180, 1999).

Therefore, there is a need to provide a means for producing glycoprotein in plants with human compatible non-immunogenic bisecting oligosaccharides.

SUMMARY OF THE INVENTION

The invention relates to expression of a mammalian N-acetylglucosaminyl-transferase III (GnTIII) enzyme in plants and its use in producing glycoproteins with bisected oligosaccharides and

increased amount of terminal G1cNAc residues. The invention further relates to a hybrid protein comprising the catalytic site of GnTIII and transmembrane domain of Golgi apparatus and/or endoplasmic reticulum (ER) protein or modified GNTIII comprising ER retention signals and its use in producing glycoproteins with oligosaccharides that lack immunogenic xylose and fucose residues.

5 In one embodiment, the present invention contemplates a plant host system comprising or expressing a mammalian UDP-Nacetylglucosamine:(β -D-mannoside β (1,4)-Nacetylglucosaminyltransferase (GnTIII) enzyme (nucleotide sequence: SEQ ID NO.: 1, Genbank I.D. number AL022312 (Dunham, I., et al., *Nature* 402:489-495, 1999); protein sequence: SEQ ID 10 NO.: 2, Genbank I.D. number Q09327), wherein said GnTIII inserts bisecting Nacetyl glucosamine (G1cNAc) residues in complex-type N-linked glycans of a glycoprotein present in said plant host system

15 In a specific embodiment of the invention, the plant host system further comprises a heterologous glycoprotein or functional fragment thereof comprising bisected oligosaccharide, particularly galactose residues. The GnTIII inserts bisecting N-G1cNAc residues onto said heterologous glycoprotein.

20 In one embodiment, the present invention contemplates to a method for obtaining a plant host system expressing a heterologous glycoprotein comprising bisecting oligosaccharides. In one embodiment, the method comprises crossing a plant expressing a heterologous glycoprotein with a plant expressing said GnTIII, harvesting progeny from said crossing and selecting a desired progeny plant expressing said heterologous glycoprotein and expressing mammalian GnTIII. Alternatively, said plant host system may be obtained by introducing into a plant or portion thereof a nucleic acid encoding said mammalian GnTIII and a nucleic acid encoding said heterologous glycoprotein and isolating a plant or portion thereof expressing said heterologous glycoprotein and expressing 25 mammalian GnTIII that is normally not present in plants. Furthermore, the invention is directed to a method for obtaining said heterologous glycoprotein from said plant comprising obtaining a plant host system using either of the procedures described above and further isolating said heterologous glycoprotein.

30 In another embodiment, it is contemplated that the plant host system of the present invention further comprises a functional mammalian enzyme providing N-glycan biosynthesis that is normally not present in plants thereby, for example, providing the capacity to extend an N-linked glycan by the addition of a galactose as described in WO 01 /21045 (herein incorporated by reference). In another embodiment, the present invention further contemplates a plant host system, wherein said plant host system comprises crossing a plant, said plant comprising a functional protein such as a transporter 35 protein or a enzyme (*e.g.*, a mammalian protein) or functional fragment thereof wherein said protein provides N-glycan biosynthesis, with a plant comprising said mammalian GnTIII. In another embodiment, the present invention contemplates harvesting the progeny from said crossing and selecting a desired progeny plant expressing said functional protein such as, for example, a transporter protein or enzyme or functional fragment thereof. In yet another embodiment of the

present invention, it is contemplated that the expressed protein provides N-glycan biosynthesis and the mammalian GnTIII. In still yet another embodiment, the present invention contemplates a plant host system, wherein a nucleic acid encoding the GnTIII and a nucleic acid encoding a functional protein (for example, a transporter or an enzyme [e.g., mammalian] or functional fragment thereof) providing N-glycan biosynthesis and isolating said plant or portion thereof expressing the functional protein or functional fragment thereof providing N-glycan biosynthesis and said mammalian GnTIII. Although the present invention is not limited to any particular theory or mechanism, it is believed that such a combination increases galactosylation of a heterologous glycoprotein. Additionally, in one embodiment, it is contemplated that GnTIII and other proteins providing N-glycosylation such as GALT can also be introduced simultaneously via one transformation vector.

In one embodiment, the present invention contemplates a plant host system comprising expressing said heterologous glycoprotein (wherein, said heterologous glycoprotein has increased galactosylation) and methods for obtaining said plant host cell system and said heterologous glycoprotein. In another embodiment, the plant host cell system may be obtained by either crossing a plant wherein the plant comprises mammalian GnTIII and a functional protein (for example, a transporter or an enzyme [e.g., mammalian] or functional fragment thereof that provides N-glycan biosynthesis not normally found in plants) with a plant comprising a heterologous glycoprotein and, then, selecting said progeny plants. In yet another embodiment, it is contemplated that said heterologous glycoprotein may be obtained by introducing nucleic acid sequences encoding 1) said GnTIII, 2) said functional protein or enzyme providing N-glycan biosynthesis not normally found in plants and 3) said heterologous glycoprotein into said plant or portion thereof and isolating said plant or portion thereof expressing said nucleic acid sequences. In another embodiment of the present invention, it is contemplated that the heterologous glycoproteins will be isolated or purified from the plant host systems.

In one embodiment of the present invention, a hybrid protein is contemplated, wherein the hybrid protein comprises 1) an isolated hybrid protein comprising a catalytic portion of mammalian GnTIII and 2) a transmembrane portion of a protein from, for example, the endoplasmic reticulum or Golgi apparatus of a eukaryotic cell. In another embodiment, the present invention also contemplates a modified mammalian GnTIII comprising a retention signal such as KDEL for retention of said GnTIII in the ER. In yet another embodiment, the present invention contemplates nucleic acid sequences encoding 1) said hybrid proteins and said modified mammalian GnTIII, 2) vectors comprising said nucleic acid sequences and 3) plant host systems comprising said sequences. In one embodiment, these hybrid proteins and modified GnTIIIs may act to relocalize GnTIII activity in the endoplasmic reticulum (ER) and/or Golgi apparatus. In another embodiment, the present invention contemplates methods for obtaining these hybrid proteins and modified GnTIII proteins by, for example, introducing sequences encoding said hybrid proteins or modified GnTIIIs into a plant or portion thereof. Although the present invention is not limited to any particular theory or mechanism, it is believed that as a result of such relocalization, bisecting GlcNAc is be introduced earlier in the N-glycan biosynthesis sequence of reactions thereby preventing subsequent enzymatic reactions and,

as a consequence, a heterologous protein expressed in a plant host system (for example, the plant host system of the present invention) will lack xylose and fucose and have increased amount of terminal GlcNAc. Accordingly, one embodiment of the present invention contemplates a method to provide a plant host system expressing a heterologous glycoprotein (said plant host system having
5 the capacity to extend an N-linked glycan with galactose) comprising crossing a plant comprising said 1) hybrid protein or said modified GnTIII with a plant comprising said heterologous protein and 2) selecting said desired progeny. In another embodiment, the present invention contemplates introducing into a plant or portion thereof a nucleic acid sequence encoding 1) said modified GnTIII or said hybrid protein and said heterologous glycoprotein and 2) isolating said plant or portion
10 thereof expressing a heterologous glycoprotein with the capacity to extend an N-linked glycan with galactose. In yet another embodiment, the present invention contemplates a method for obtaining said desired heterologous glycoprotein, said method comprising isolating said glycoprotein from said plant or portion thereof.

In one embodiment, the present invention contemplates that the plant-derived glycoprotein or
15 functional fragment thereof may be used for the production of a pharmaceutical composition (for example, an antibody, a hormone, a vaccine antigen, an enzyme, or the like). In another embodiment, the present invention contemplates a pharmaceutical composition comprising a glycoprotein or functional fragment thereof is now also provided.

In one embodiment, the present invention contemplates variants or mutants of GnTIII. The
20 terms "variant" and "mutant" when used in reference to a polypeptide refer to an amino acid sequence that differs by one or more amino acids from another, usually related, polypeptide. In another embodiment, the present invention contemplates variants that have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties. One type of conservative amino acid substitutions refers to the interchangeability of residues having similar side
25 chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino
30 acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine (V) -leucine (L) -isoleucine (I), phenylalanine (F) -tyrosine (Y), lysine (K) -arginine (R), alanine (A) -valine (V), and asparagine (N) -glutamine (Q).

In yet another embodiment, the present invention contemplates variants that have "non-conservative" changes (*e.g.*, replacement of a glycine with a tryptophan). Similar minor variations
35 may also include amino acid deletions or insertions (*i.e.*, additions), or both. Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing biological activity may be found using computer programs well known in the art, for example, DNAStar software. Variants can be tested in functional assays. For both conservative and non-conservative variants, preferred variants have less than 10 %, preferably less than 5 % and,

still more preferably, less than 2 % changes (whether substitutions, deletions, and so on).

In one embodiment, the present invention contemplates a plant host (cell) system, comprising a mammalian UDP-N-acetylglucosamine: β -D mannoside β (1,4)-N-acetylglucosaminyltransferase (GnTIII) enzyme (or portion or variant thereof, wherein said GnTIII inserts bisecting N-acetyl glucosamine (G1cNAc) residues in complex-type N-linked glycans of a glycoprotein present in said plant host system). In another embodiment, the present invention contemplates the plant host, wherein said GnTIII is a human GnTIII. In yet another embodiment, the present invention contemplates the plant host system, wherein said system is a portion of a plant. In yet another embodiment, the present invention contemplates the plant host system, wherein said system is a portion of a plant selected from the group consisting of a cell, leaf, embryo, callus, stem, pericarp, protoplast, root, tuber, kernel, endosperm and embryo. In yet another embodiment, the present invention contemplates the plant host system, wherein said system is a whole plant. In yet another embodiment, the present invention contemplates the plant host system, further comprising a heterologous glycoprotein (or functional fragment thereof). In yet another embodiment, the present invention contemplates the plant host system, wherein said heterologous glycoprotein protein comprises an antibody, or fragment (e.g. Fc, Fv, Fab, Fab₂) thereof. In yet another embodiment, the present invention contemplates the plant host system, wherein said heterologous glycoprotein or functional fragment thereof comprises bisected oligosaccharides. In yet another embodiment, the present invention contemplates the plant host system, wherein said heterologous glycoprotein (or functional fragment thereof) comprises bisected glycans with galactose residues. In yet another embodiment, the present invention contemplates the plant host system, wherein said plant is a tobacco plant. In yet another embodiment, the present invention contemplates the plant host system, which further comprises a functional protein selected from a group consisting of a transporter or a (mammalian) enzyme (or functional fragment thereof) providing N-glycan biosynthesis. In yet another embodiment, the present invention contemplates the plant host system, wherein said enzyme is a (human) β -1,4 galactosyltransferase. In yet another embodiment, the present invention contemplates the plant host system, which further comprises a heterologous glycoprotein, having an increased number of galactose residues. In yet another embodiment, the present invention contemplates a plant host system comprising a nucleic acid sequence encoding a mammalian GnTIII protein. In yet another embodiment, the present invention contemplates a plant host system comprising a vector comprising a nucleic acid sequence encoding a mammalian GnTIII protein. In yet another embodiment, the present invention contemplates the plant host, which further comprises a nucleic acid sequence encoding a functional protein selected from a group consisting of a transporter or a (mammalian) enzyme (or functional fragment thereof) providing N-glycan biosynthesis.

In one embodiment, the present invention contemplates a method (for obtaining a plant host system expressing a heterologous glycoprotein having bisected oligosaccharides) comprising a) crossing a plant expressing a heterologous glycoprotein with a, b) harvesting progeny from said crossing and c) selecting a desired progeny plant (expressing said heterologous glycoprotein and

expressing a mammalian GnTIII that is normally not present in plants). In another embodiment, the present invention contemplates this method, wherein said desired progeny plant expresses said heterologous glycoprotein protein having bisected oligosaccharides. In yet another embodiment, the present invention contemplates this method, wherein said plant host system is a transgenic plant.

5 In one embodiment, the present invention contemplates a method for obtaining a heterologous glycoprotein having bisected oligosaccharides comprising a) introducing a nucleic acid sequence encoding GnTIII that is normally not present in plant into a plant host system and a nucleic acid sequence encoding a heterologous glycoprotein and b) isolating said heterologous glycoprotein. In another embodiment, the present invention contemplates this method, wherein said nucleic acid sequences are introduced into a plant cell and said plant cell is regenerated into a plant. In yet another embodiment, the present invention contemplates the same method, wherein said nucleic acid sequences are introduced into a plant host system by transforming said plant host system with a vector comprising a nucleic acid sequence encoding GnTIII that is normally not present in plant into a plant and a nucleic acid sequence encoding a heterologous glycoprotein. In yet another embodiment, the present invention contemplates the method, wherein said nucleic acid sequences are introduced into a plant host system by transforming said plant host system with a vector comprising a nucleic acid sequence encoding GnTIII that is normally not present in plant into a plant and a nucleic acid sequence encoding a heterologous glycoprotein. In yet another embodiment, the present invention contemplates the method, wherein said nucleic acid sequences are introduced into a plant host system by transforming said plant with a vector comprising a nucleic acid sequence encoding GnTIII that is normally not present in plant into a plant host system and vector comprising a nucleic acid sequence encoding a heterologous glycoprotein. In yet another embodiment, the present invention contemplates a method for obtaining a heterologous glycoprotein having bisected oligosaccharides comprising cultivating the regenerated plant.

25 In one embodiment, the present invention contemplates a method for obtaining a desired glycoprotein (or functional fragment thereof) comprising a) cultivating the plant host system (until said plant has reached a harvestable stage) and b) harvesting said plant (and fractionating to obtain fractionated plant material and c) at least partly isolating said glycoprotein from said fractionated plant material). In another embodiment, the present invention contemplates a plant obtainable by the contemplated method.

30 In one embodiment, the present invention contemplates A method for obtaining a plant host system comprising a functional protein selected from a group consisting of a transporter or a (mammalian) enzyme or functional fragment thereof providing N-glycan biosynthesis and a mammalian GnTIII comprising crossing a plant comprising a functional protein such as a transporter or a (mammalian) enzyme or functional fragment thereof providing N-glycan biosynthesis with a plant according to Claim 5, harvesting progeny from said crossing and selecting a desired progeny plant expressing said functional protein such as a transporter or a (mammalian) enzyme or functional fragment thereof providing N-glycan biosynthesis and said mammalian GnTIII. In another embodiment, the present invention contemplates a transgenic plant obtained according to the

contemplated.

In one embodiment, the present invention contemplates a method for increasing galactosylation of a heterologous glycoprotein expressed in a plant host system comprising introducing a nucleic acid sequence encoding GnTIII and a sequence selected from a group 5 consisting sequences that encode a transporter or a (mammalian) enzyme or functional fragment not normally present in a plant into said plant host system expressing said heterologous glycoprotein and isolating said glycoprotein.

In one embodiment, the present invention contemplates a plant derived glycoprotein comprising bisected oligosaccharides.

10 In one embodiment, the present invention contemplates the use of a plant host system contemplated by the present invention to produce a desired glycoprotein or functional fragment thereof. In another embodiment, the present invention contemplates that said glycoprotein or functional fragment thereof comprises bisected oligosaccharides. In yet another embodiment, the present invention contemplates a plant-derived glycoprotein or functional fragment thereof obtained 15 by a method contemplated by the present invention. In yet another embodiment, the present invention contemplates a glycoprotein or functional fragment thereof contemplated by the invention for the production of a pharmaceutical composition. In yet another embodiment, the present invention contemplates a composition comprising a glycoprotein or functional fragment thereof as contemplated by the present invention.

20 In one embodiment, the present invention contemplates an isolated hybrid protein comprising an active site of GnTIII and a transmembrane region of a protein, said protein residing in endoplasmic reticulum or Golgi apparatus of a eukaryotic cell. In another embodiment, the present invention contemplates the protein of the present invention, wherein said protein residing in endoplasmic reticulum or Golgi apparatus of a eukaryotic cell is an enzyme. In yet another 25 embodiment, the present invention contemplates the protein accordof the present invention, wherein said protein residing in endoplasmic reticulum or Golgi apparatus of a eukaryotic cell is a glycosyltransferase. In yet another embodiment, the present invention contemplates the protein of the present invention, wherein said protein residing in endoplasmic reticulum or Golgi apparatus of a eukaryotic cell is a glycosyltransferase selected from the group consisting of a mannosidaseI, 30 mannosidaseII, GnTI, GnTII, Xy1T and FucT. In yet another embodiment, the present invention contemplates the protein accorof the present invention, wherein said protein residing in endoplasmic reticulum or Golgi apparatus of a eukaryotic cell is a plant protein. In yet another embodiment, the present invention contemplates an isolated nucleic acid sequence encoding the protein of the present invention. In yet another embodiment, the present invention contemplates a vector comprising the 35 isolated nucleic acid sequence of the present invention. In yet another embodiment, the present invention contemplates a plant comprising the isolated nucleic acid sequence of the present invention. In yet another embodiment, the present invention contemplates the plant(s) of the present invention which further comprises a nucleic acid sequence encoding a heterologous glycoprotein.

In one embodiment, the present invention contemplates a method (for providing a transgenic

plant capable of expressing a heterologous glycoprotein with the capacity to extend an N-linked glycan with galactose) comprising a) crossing a transgenic plant with a plant of the present invention, b) harvesting progeny from said crossing and c) selecting a desired progeny plant (expressing said recombinant protein and expressing a functional (mammalian) enzyme involved in (mammalian) N-glycan biosynthesis that is normally not present in plants).

In one embodiment, the present invention contemplates a method for providing a transgenic plant capable of expressing a heterologous glycoprotein with the capacity to extend an N-linked glycan with galactose comprising introducing the nucleic acid sequence of the present invention and a nucleic acid sequence encoding said heterologous glycoprotein.

10 In one embodiment, the present invention contemplates a method, comprising: a) providing: i) a plant cell, and ii) an expression vector comprising nucleic acid encoding a GNTIII enzyme; and b) introducing said expression vector into said plant cell under conditions such that said enzyme is expressed. In another embodiment, the present invention contemplates the method, wherein said nucleic acid encoding a GNTIII comprises the nucleic acid sequence of SEQ ID NO:1.

15 In one embodiment, the present invention contemplates a method, comprising: a) providing: i) a plant cell, ii) a first expression vector comprising nucleic acid encoding a GNTIII enzyme, and iii) a second expression vector comprising nucleic acid encoding a heterologous glycoprotein; and b) introducing said first and second expression vectors into said plant cell under conditions such that said hybrid enzyme and said heterologous protein are expressed. In another embodiment, the present 20 invention contemplates the method, wherein said heterologous protein is an antibody or antibody fragment.

In one embodiment, the present invention contemplates A method, comprising: a) providing: i) a first plant comprising a first expression vector, said first vector comprising nucleic acid encoding a GNTIII enzyme, and ii) a second plant comprising a second expression vector, said 25 second vector comprising nucleic acid encoding a heterologous protein; and b) crossing said first plant and said second plant to produce progeny expressing said hybrid enzyme and said heterologous protein.

30 In one embodiment, the present invention contemplates a plant, comprising first and second expression vectors, said first vector comprising nucleic acid encoding a GNTIII enzyme, said second vector comprising nucleic acid encoding a heterologous protein. In another embodiment, the present invention contemplates the, wherein said heterologous protein is an antibody or antibody fragment.

BRIEF DESCRIPTION OF THE FIGURES

Figures 1A and 1B show MALDI-TOF mass spectra of (A) N-linked glycans isolated from 35 leaves of control tobacco plant and (B) N-linked glycans isolated from leaves of selected GnTM-17 tobacco plant transformed with human GnTIII. See, Table 1 for structures.

Figure 2 shows processing of high mannose type glycan (M9) to complex type glycans under the subsequent action of ManI, GnTI, ManII, and GnTII. It is also indicated what glycan structures the action of Ga1T and/or GnTIII at different points

in chain of reactions would lead. The reactions catalyzed by fucosyltransferases and xylosyltransferases are not indicated. Core G1cNAc (Gn) is not indicated. Gn = G1cNAc, Gn^b = bisecting G1cNAc, G = galactose and M = mannose.

Figures 3A and 3B show (A) the T-DNA construct carrying the genes encoding glycan modifying enzymes to produce efficiently galactosylated bisected glycans that are devoid of immunogenic xylose and fucose and (B) the T-DNA construct carrying antibody light chain and heavy chain genes. TmXyl = transmembrane domain of xylosyltransferase, TmGnTI = transmembrane domain of GnT, P = promoter, R = selection marker, L = antibody light chain and H = antibody heavy chain.

Figures 4A and 4B show the nucleotide sequence (SEQ ID NO: 1, underlined portion of Figure 4(A) the protein sequence (SEQ ID NO: 2, underlined portion of Figure 4B) of *GnTIII* including a c-myc tag. Residues that can undergo conservative amino acid substitutions are defined in the DEFINITIONS section.

Figure 5A and 5B show a (A) map of the plasmid pDAB4005 and (B) the nucleotide sequence of the plasmid pDAB4005 (SEQ ID NO: 8).

Figure 6A and 6B show a (A) map of the plasmid pDAB7119 and (B) the nucleotide sequence of the plasmid pDAB7119 (SEQ ID NO: 9) including splice sites.

Figure 7A and 7B show a (A) map of the plasmid pDAB8504 and (B) the nucleotide sequence of the plasmid pDAB8504 (SEQ ID NO: 10).

Figure 8A and 8B show a (A) map of the plasmid pDAB7113 and (B) the nucleotide sequence of the plasmid pDAB7113 (SEQ ID NO: 11) including splice sites.

Figures 9A and 9B show MALDI-TOF mass spectra of glycoproteins from control and *GnTIII* corn. Comparison of mass spectra of N-glycans of glycoproteins isolated from calli of (A) control corn and of (B) selected *GnTIII*-corn. *GnTIII* corn was obtained through transformation with human *GnTIII* gene sequence and selection was performed by lectin blotting using E-PHA. See Table 3 for an annotation of the data contained in Figures 9A and 9B.

Figure 10 shows the full nucleotide sequence of *GntIII* without a c-myc tag (SEQ ID NO: 7).

Figure 11 shows a MALDI-TOF mass spectra of glycoproteins from control and *GnTIII* corn-2. See, Table 4 for structures and abbreviations.

Figure 12 shows a representative blot of samples of transgenic maize callus for altered lectin binding due to expression of the *GntIII* gene.

Figure 13 shows a representative blot of samples of transgenic maize callus for c-myc epitope expression.

Figures 14 A and 14B show a MALDI-TOF mass spectra of glycoproteins from (A) control and (B) *GnTIII* corn plants.

DEFINITIONS

The terms "protein" and "polypeptide" refer to compounds comprising amino acids joined via peptide bonds and are used interchangeably. A "protein" or "polypeptide" encoded by a gene is not

limited to the amino acid sequence encoded by the gene, but includes post-translational modifications of the protein.

The term "glycoprotein" refers to proteins with covalently attached sugar units, either bonded via the OH group of serine or threonine (O glycosylated) or through the amide NH₂ of asparagine (N glycosylated). "Glycoprotein" may include, but is not limited to, for example, most secreted proteins (serum albumin is the major exception) and proteins exposed at the outer surface of the plasma membrane. Sugar residues found include, but are not limited to: mannose, N acetyl glucosamine, N acetyl galactosamine, galactose, fucose and sialic acid.

Where the term "amino acid sequence" is recited herein to refer to an amino acid sequence of a protein molecule, "amino acid sequence" and like terms, such as "polypeptide" or "protein" are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule. Furthermore, an "amino acid sequence" can be deduced from the nucleic acid sequence encoding the protein.

The term "portion" when used in reference to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from four amino acid residues to the entire amino sequence minus one amino acid.

The term "chimera" when used in reference to a polypeptide refers to the expression product of two or more coding sequences obtained from different genes, that have been cloned together and that, after translation, act as a single polypeptide sequence. Chimeric polypeptides are also referred to as "hybrid" polypeptides. The coding sequences includes those obtained from the same or from different species of organisms.

The term "fusion" when used in reference to a polypeptide refers to a chimeric protein containing a protein of interest joined to an exogenous protein fragment (the fusion partner). The fusion partner may serve various functions, including enhancement of solubility of the polypeptide of interest, as well as providing an "affinity tag" to allow purification of the recombinant fusion polypeptide from a host cell or from a supernatant or from both. If desired, the fusion partner may be removed from the protein of interest after or during purification.

The term "homolog" or "homologous" when used in reference to a polypeptide refers to a high degree of sequence identity between two polypeptides, or to a high degree of similarity between the three-dimensional structure or to a high degree of similarity between the active site and the mechanism of action. In a preferred embodiment, a homolog has a greater than 60 % sequence identity, and more preferably greater than 75% sequence identity, and still more preferably greater than 90 % sequence identity, with a reference sequence.

As applied to polypeptides, the term "substantial identity" means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 80 percent sequence identity, preferably at least 90 percent sequence identity, more preferably at least 95 percent sequence identity or more (e.g., 99 percent sequence identity). Preferably, residue positions which are not identical differ by conservative amino acid substitutions.

The terms "variant" and "mutant" when used in reference to a polypeptide refer to an amino

acid sequence that differs by one or more amino acids from another, usually related polypeptide. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties. One type of conservative amino acid substitutions refers to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine (V) - leucine (L) -isoleucine (I), phenylalanine (F) -tyrosine (Y), lysine (K) -arginine (R), alanine (A) - valine (V), and asparagine (N) -glutamine (Q). More rarely, a variant may have "non-conservative" changes (*e.g.*, replacement of a glycine with a tryptophan). Similar minor variations may also include amino acid deletions or insertions (*i.e.*, additions), or both. Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing biological activity may be found using computer programs well known in the art, for example, DNAStar software. Variants can be tested in functional assays. Preferred variants have less than 10 %, and preferably less than 5 %, and still more preferably less than 2 % changes (whether substitutions, deletions, and so on).

The term "domain" when used in reference to a polypeptide refers to a subsection of the polypeptide which possesses a unique structural and/or functional characteristic; typically, this characteristic is similar across diverse polypeptides. The subsection typically comprises contiguous amino acids, although it may also comprise amino acids which act in concert or which are in close proximity due to folding or other configurations.

The term "gene" refers to a nucleic acid (*e.g.*, DNA or RNA) sequence that comprises coding sequences necessary for the production of an RNA, or a polypeptide or its precursor (*e.g.*, proinsulin). A functional polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence as long as the desired activity or functional properties (*e.g.*, enzymatic activity, ligand binding, signal transduction, etc.) of the polypeptide are retained. The term "portion" when used in reference to a gene refers to fragments of that gene. The fragments may range in size from a few nucleotides (*e.g.*, ten nucleotides) to the entire gene sequence minus one nucleotide. Thus, "a nucleotide comprising at least a portion of a gene" may comprise fragments of the gene or the entire gene.

The term "gene" also encompasses the coding regions of a structural gene and includes sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb on either end such that the gene corresponds to the length of the full-length mRNA. The sequences which are located 5' of the coding region and which are present on the mRNA are referred to as 5' non-translated sequences. The sequences which are located 3' or downstream of the coding region and which are present on the mRNA are referred to as 3' non-translated sequences. The term "gene"

encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene which are transcribed into nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or
5 "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' end of the sequences which are present on the RNA transcript. These
10 sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers which control or influence the transcription of the gene. The 3' flanking region may contain sequences which direct the termination of transcription, posttranscriptional cleavage and polyadenylation.

15 The term "heterologous" when used in reference to a gene refers to a gene encoding a factor that is not in its natural environment (*i.e.*, has been altered by the hand of man). For example, a heterologous gene includes a gene from one species introduced into another species. A heterologous gene also includes a gene native to an organism that has been altered in some way (*e.g.*, mutated, added in multiple copies, linked to a non-native promoter or enhancer sequence, etc.). Heterologous
20 genes may comprise gene sequences that comprise cDNA forms of a gene; the cDNA sequences may be expressed in either a sense (to produce mRNA) or anti-sense orientation (to produce an anti-sense RNA transcript that is complementary to the mRNA transcript). Heterologous genes are distinguished from endogenous genes in that the heterologous gene sequences are typically joined to nucleotide sequences comprising regulatory elements such as promoters that are not found naturally
25 associated with the gene for the protein encoded by the heterologous gene or with gene sequences in the chromosome, or are associated with portions of the chromosome not found in nature (*e.g.*, genes expressed in loci where the gene is not normally expressed).

30 A "heterologous glycoprotein" is a glycoprotein originating from a species other than the plant host system. The glycoprotein may include but is not limited to antibodies, hormones, growth factors, and growth factor receptors, antigens, cytokines and blood products.

35 A "plant host system" may include, but is not limited to, a plant or portion thereof which includes, but is not limited to, a plant cell, plant organ and/or plant tissue. The plant may be a monocotyledon (monocot) which is a flowering plant whose embryos have one cotyledon or seed leaf and includes but is not limited to lilies, grasses, corn (*Zea mays*), rice, grains including oats, wheat and barley, orchids, irises, onions and palms. Alternatively, the plant may be a dicotyledon (dicot) which includes, but is not limited to, tobacco (*Nicotiana*), tomatoes, potatoes, legumes (*e.g.*, alfalfa and soybeans), roses, daises, cacti, violets and duckweed. The plant may also be a moss which includes, but is not limited to, *Physcomitrella patens*. The invention is further directed to a method for obtaining said bisected G1cNAc in a plant host system by introducing a nucleic acid

encoding said GnTIII into a plant or portion thereof and expressing said GnTIII and isolating said plant or portion thereof expressing said GnTIII.

The term "nucleotide sequence of interest" or "nucleic acid sequence of interest" refers to any nucleotide sequence (*e.g.*, RNA or DNA), the manipulation of which may be deemed desirable for 5 any reason (*e.g.*, treat disease, confer improved qualities, *etc.*), by one of ordinary skill in the art. Such nucleotide sequences include, but are not limited to, coding sequences of structural genes (*e.g.*, reporter genes, selection marker genes, oncogenes, drug resistance genes, growth factors, *etc.*), and non-coding regulatory sequences which do not encode an mRNA or protein product (*e.g.*, promoter 10 sequence, polyadenylation sequence, termination sequence, enhancer sequence and other like sequences). The present invention contemplates host cells expressing a heterologous protein encoded by a nucleotide sequence of interest along with one or more hybrid enzymes.

The term "structural" when used in reference to a gene or to a nucleotide or nucleic acid sequence refers to a gene or a nucleotide or nucleic acid sequence whose ultimate expression product is a protein (such as an enzyme or a structural protein), an rRNA, an sRNA, a tRNA, etc.

15 The terms "oligonucleotide" or "polynucleotide" or "nucleotide" or "nucleic acid" refer to a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than three, and usually more than ten. The exact size will depend on many factors, which in turn depends on the ultimate function or use of the oligonucleotide. The oligonucleotide may be generated in any manner, including chemical synthesis, DNA replication, reverse transcription, or a combination 20 thereof.

25 The terms "an oligonucleotide having a nucleotide sequence encoding a gene" or "a nucleic acid sequence encoding" a specified polypeptide refer to a nucleic acid sequence comprising the coding region of a gene or in other words the nucleic acid sequence which encodes a gene product. The coding region may be present in either a cDNA, genomic DNA or RNA form. When present in a DNA form, the oligonucleotide may be single-stranded (*i.e.*, the sense strand) or double-stranded. Suitable control elements such as enhancers/promoters, splice junctions, polyadenylation signals, *etc.* 30 may be placed in close proximity to the coding region of the gene if needed to permit proper initiation of transcription and/or correct processing of the primary RNA transcript. Alternatively, the coding region utilized in the expression vectors of the present invention may contain endogenous enhancers/promoters, splice junctions, intervening sequences, polyadenylation signals, *etc.* or a combination of both endogenous and exogenous control elements.

35 The term "recombinant" when made in reference to a nucleic acid molecule refers to a nucleic acid molecule which is comprised of segments of nucleic acid joined together by means of molecular biological techniques. The term "recombinant" when made in reference to a protein or a polypeptide refers to a protein molecule which is expressed using a recombinant nucleic acid molecule.

As used herein, the terms "complementary" or "complementarity" are used in reference to nucleotide sequences related by the base-pairing rules. For example, the sequence 5'-AGT-3' is complementary to the sequence 5'-ACT-3'. Complementarity can be "partial" or "total." "Partial"

complementarity is where one or more nucleic acid bases is not matched according to the base pairing rules. "Total" or "complete" complementarity between nucleic acids is where each and every nucleic acid base is matched with another base under the base pairing rules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength 5 of hybridization between nucleic acid strands.

A "complement" of a nucleic acid sequence as used herein refers to a nucleotide sequence whose nucleic acids show total complementarity to the nucleic acids of the nucleic acid sequence. For example, the present invention contemplates the complements of SEQ ID NO: 1.

The term "homology" when used in relation to nucleic acids refers to a degree of

10 complementarity. There may be partial homology (*i.e.*, partial identity) or complete homology (*i.e.*, complete identity). A partially complementary sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid and is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay
15 (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe (*i.e.*, an oligonucleotide which is capable of hybridizing to another oligonucleotide of interest) will compete for and inhibit the binding (*i.e.*, the hybridization) of a completely homologous sequence to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted;
20 low stringency conditions require that the binding of two sequences to one another be a specific (*i.e.*, selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of complementarity (*e.g.*, less than about 30 % identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

25 When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described *infra*.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially 30 homologous" refers to any probe which can hybridize to the single-stranded nucleic acid sequence under conditions of low stringency as described *infra*.

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence," "sequence identity," "percentage of sequence identity" and "substantial identity." A "reference sequence" is a defined sequence used as a basis for a sequence 35 comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA sequence given in a sequence listing or may comprise a complete gene sequence. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (*i.e.*, a portion of the complete polynucleotide sequence) that is similar between

the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window," as used herein,

5 refers to a conceptual segment of at least 20 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman [Smith and Waterman, *Adv. Appl. Math.* 2: 10 482 (1981)] by the homology alignment algorithm of Needleman and Wunsch [Needleman and Wunsch, *J. Mol. Biol.* 48:443 (1970)], by the search for similarity method of Pearson and Lipman [Pearson and Lipman, *Proc. Natl. Acad. Sci. (U.S.A.)* 85:2444 (1988)], by computerized 15 implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by inspection, and the best alignment (*i.e.*, resulting in the highest percentage of homology over the comparison window) generated by the various methods is selected. The term "sequence identity" means that two polynucleotide sequences are identical (*i.e.*, on a nucleotide-by-nucleotide 20 basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (*i.e.*, the window size), and multiplying the 25 result by 100 to yield the percentage of sequence identity. The terms "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 85 percent sequence identity, preferably at least 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25-50 30 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. The reference sequence may be a subset of a larger sequence, for example, as a segment of the full-length sequences of the compositions claimed in the present invention.

35 The term "hybridization" refers to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids. A single molecule that contains pairing of complementary nucleic acids within its structure is said to be

"self-hybridized."

The term " T_m " refers to the "melting temperature" of a nucleic acid. The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\% G + C)$, when a nucleic acid is in aqueous solution at 1 M NaCl (See e.g., Anderson and Young, Quantitative Filter Hybridization, in *Nucleic Acid Hybridization* [1985]). Other references include more sophisticated computations that take structural as well as sequence characteristics into account for the calculation of T_m .

The term "stringency" refers to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. With "high stringency" conditions, nucleic acid base pairing will occur only between nucleic acid fragments that have a high frequency of complementary base sequences. Thus, conditions of "low" stringency are often required with nucleic acids that are derived from organisms that are genetically diverse, as the frequency of complementary sequences is usually less.

"Low stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42 °C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH₂PO₄(H₂O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5X Denhardt's reagent [50X Denhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)] and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5X SSPE, 0.1% SDS at 42 °C when a probe of about 500 nucleotides in length is employed.

"Medium stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42 °C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH₂PO₄(H₂O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5 % SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 1.0X SSPE, 1.0 % SDS at 42 °C when a probe of about 500 nucleotides in length is employed.

"High stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42 °C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH₂PO₄(H₂O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5 % SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1X SSPE, 1.0 % SDS at 42 °C when a probe of about 500 nucleotides in length is employed.

It is well known that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may

be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, the art knows conditions that promote hybridization under conditions of high stringency (*e.g.*, increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.).

5 Additionally, the term "equivalent," when made in reference to a hybridization condition as it relates to a hybridization condition of interest, means that the hybridization condition and the hybridization condition of interest result in hybridization of nucleic acid sequences which have the same range of percent (%) homology. For example, if a hybridization condition of interest results in hybridization of a first nucleic acid sequence with other nucleic acid sequences that have from 50 %
10 to 70 % homology to the first nucleic acid sequence, then another hybridization condition is said to be equivalent to the hybridization condition of interest if this other hybridization condition also results in hybridization of the first nucleic acid sequence with the other nucleic acid sequences that have from 50 % to 70 % homology to the first nucleic acid sequence.

When used in reference to nucleic acid hybridization the art knows well that numerous
15 equivalent conditions may be employed to comprise either low or high stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (*e.g.*, the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of either
20 low or high stringency hybridization different from, but equivalent to, the above-listed conditions.

The term "wild-type" when made in reference to a gene refers to a gene that has the characteristics of a gene isolated from a naturally occurring source. The term "wild-type" when made in reference to a gene product refers to a gene product that has the characteristics of a gene product isolated from a naturally occurring source. The term "naturally-occurring" as applied to an
25 object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally- occurring. A wild-type gene is frequently that gene which is most frequently observed in a population and is thus arbitrarily designated the "normal" or "wild-type" form of the gene. In
30 contrast, the term "modified" or "mutant" when made in reference to a gene or to a gene product refers, respectively, to a gene or to a gene product which displays modifications in sequence and/or functional properties (*i.e.*, altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally-occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

35 Thus, the terms "variant" and "mutant" when used in reference to a nucleotide sequence refer to an nucleic acid sequence that differs by one or more nucleotides from another, usually related nucleotide acid sequence. A "variation" is a difference between two different nucleotide sequences; typically, one sequence is a reference sequence.

The term "polymorphic locus" refers to a genetic locus present in a population that shows

variation between members of the population (*i.e.*, the most common allele has a frequency of less than 0.95). Thus, "polymorphism" refers to the existence of a character in two or more variant forms in a population. A "single nucleotide polymorphism" (or SNP) refers a genetic locus of a single base which may be occupied by one of at least two different nucleotides. In contrast, a "monomorphic" 5 locus" refers to a genetic locus at which little or no variations are seen between members of the population (generally taken to be a locus at which the most common allele exceeds a frequency of 0.95 in the gene pool of the population).

A "frameshift mutation" refers to a mutation in a nucleotide sequence, usually resulting from insertion or deletion of a single nucleotide (or two or four nucleotides) which results in a change in 10 the correct reading frame of a structural DNA sequence encoding a protein. The altered reading frame usually results in the translated amino-acid sequence being changed or truncated.

A "splice mutation" refers to any mutation that affects gene expression by affecting correct RNA splicing. Splicing mutation may be due to mutations at intron-exon boundaries which alter splice sites.

15 The term "detection assay" refers to an assay for detecting the presence or absence of a sequence or a variant nucleic acid sequence (*e.g.*, mutation or polymorphism in a given allele of a particular gene, as *e.g.*, *GnTIII* gene, SEQ ID NO: 1, Figure 4A), or for detecting the presence or absence of a particular protein (*e.g.*, *GnTIII*, SEQ ID NO: 2, Figure 4B) or the structure or activity or effect of a particular protein (*e.g.*, *GnTIII* activity), for detecting glycosylation moieties on a 20 particular protein (*e.g.*, N-linked glycans) or for detecting the presence or absence of a variant of a particular protein.

The term "antisense" refers to a deoxyribonucleotide sequence whose sequence of deoxyribonucleotide residues is in reverse 5' to 3' orientation in relation to the sequence of deoxyribonucleotide residues in a sense strand of a DNA duplex. A "sense strand" of a DNA duplex 25 refers to a strand in a DNA duplex which is transcribed by a cell in its natural state into a "sense mRNA." Thus an "antisense" sequence is a sequence having the same sequence as the non-coding strand in a DNA duplex. The term "antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression 30 of a target gene by interfering with the processing, transport and/or translation of its primary transcript or mRNA. The complementarity of an antisense RNA may be with any part of the specific gene transcript, *i.e.*, at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. In addition, as used herein, antisense RNA may contain regions of ribozyme sequences 35 that increase the efficacy of antisense RNA to block gene expression. "Ribozyme" refers to a catalytic RNA and includes sequence-specific endoribonucleases. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of preventing the expression of the target protein.

"Amplification" is a special case of nucleic acid replication involving template specificity. It is to be contrasted with non-specific template replication (*i.e.*, replication that is template-dependent but not dependent on a specific template). Template specificity is here distinguished from fidelity of

replication (*i.e.*, synthesis of the proper polynucleotide sequence) and nucleotide (ribo- or deoxyribo-) specificity. Template specificity is frequently described in terms of "target" specificity. Target sequences are "targets" in the sense that they are sought to be sorted out from other nucleic acid. Amplification techniques have been designed primarily for this sorting out.

- 5 Template specificity is achieved in most amplification techniques by the choice of enzyme. Amplification enzymes are enzymes that, under conditions they are used, will process only specific sequences of nucleic acid in a heterogeneous mixture of nucleic acid. For example, in the case of Q β replicase, MDV-1 RNA is the specific template for the replicase (Kacian *et al.*, Proc. Natl. Acad. Sci. USA, 69:3038, 1972). Other nucleic acid will not be replicated by this amplification enzyme.
- 10 Similarly, in the case of T7 RNA polymerase, this amplification enzyme has a stringent specificity for its own promoters (Chamberlain *et al.*, Nature, 228:227, 1970). In the case of T4 DNA ligase, the enzyme will not ligate the two oligonucleotides or polynucleotides, where there is a mismatch between the oligonucleotide or polynucleotide substrate and the template at the ligation junction (Wu and Wallace, Genomics, 4:560, 1989). Finally, *Taq* and *Pfu* polymerases, by virtue of their ability to
- 15 function at high temperature, are found to display high specificity for the sequences bounded and thus defined by the primers; the high temperature results in thermodynamic conditions that favor primer hybridization with the target sequences and not hybridization with non-target sequences (H.A. Erlich (ed.), *PCR Technology*, Stockton Press, 1989).

- 20 The term "amplifiable nucleic acid" refers to nucleic acids that may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" will usually comprise "sample template."

- 25 The term "sample template" refers to nucleic acid originating from a sample that is analyzed for the presence of "target" (defined below). In contrast, "background template" is used in reference to nucleic acid other than sample template that may or may not be present in a sample. Background template is most often inadvertent. It may be the result of carryover, or it may be due to the presence of nucleic acid contaminants sought to be purified away from the sample. For example, nucleic acids from organisms other than those to be detected may be present as background in a test sample.

- 30 The term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (*i.e.*, in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

35 The term "probe" refers to an oligonucleotide (*i.e.*, a sequence of nucleotides), whether

occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, that is capable of hybridizing to another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used in the present invention will be 5 labeled with any "reporter molecule," so that is detectable in any detection system, including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

The term "target," when used in reference to the polymerase chain reaction, refers to the 10 region of nucleic acid bounded by the primers used for polymerase chain reaction. Thus, the "target" is sought to be sorted out from other nucleic acid sequences. A "segment" is defined as a region of nucleic acid within the target sequence.

The term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis U.S. Patent Nos. 4,683,195, 4,683,202, and 4,965,188, that describe a method for increasing the 15 concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect 20 amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing, and polymerase extension can be repeated many times (*i.e.*, denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of 25 an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because 30 the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified."

With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (*e.g.*, hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of ³²P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the 35 amplified segment). In addition to genomic DNA, any oligonucleotide or polynucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications.

The terms "PCR product," "PCR fragment," and "amplification product" refer to the resultant

mixture of compounds after two or more cycles of the PCR steps of denaturation, annealing and extension are complete. These terms encompass the case where there has been amplification of one or more segments of one or more target sequences.

The term "amplification reagents" refers to those reagents (deoxyribonucleotide

5 triphosphates, buffer, etc.), needed for amplification except for primers, nucleic acid template, and the amplification enzyme. Typically, amplification reagents along with other reaction components are placed and contained in a reaction vessel (test tube, microwell, etc.).

The term "reverse-transcriptase" or "RT-PCR" refers to a type of PCR where the starting material is mRNA. The starting mRNA is enzymatically converted to complementary DNA or
10 "cDNA" using a reverse transcriptase enzyme. The cDNA is then used as a "template" for a "PCR" reaction

The term "gene expression" refers to the process of converting genetic information encoded in a gene into RNA (*e.g.*, mRNA, rRNA, tRNA, or snRNA) through "transcription" of the gene (*i.e.*, via the enzymatic action of an RNA polymerase), and into protein, through "translation" of mRNA.
15 Gene expression can be regulated at many stages in the process. "Up-regulation" or "activation" refers to regulation that increases the production of gene expression products (*i.e.*, RNA or protein), while "down-regulation" or "repression" refers to regulation that decrease production. Molecules (*e.g.*, transcription factors) that are involved in up-regulation or down-regulation are often called "activators" and "repressors," respectively.

20 The terms "in operable combination," "in operable order" and "operably linked" refer to the linkage of nucleic acid sequences in such a manner that a nucleic acid molecule capable of directing the transcription of a given gene and/or the synthesis of a desired protein molecule is produced. The term also refers to the linkage of amino acid sequences in such a manner so that a functional protein is produced.

25 The term "regulatory element" refers to a genetic element which controls some aspect of the expression of nucleic acid sequences. For example, a promoter is a regulatory element which facilitates the initiation of transcription of an operably linked coding region. Other regulatory elements are splicing signals, polyadenylation signals, termination signals, *etc.*

Transcriptional control signals in eukaryotes comprise "promoter" and "enhancer" elements.
30 Promoters and enhancers consist of short arrays of DNA sequences that interact specifically with cellular proteins involved in transcription (Maniatis, *et al.*, Science 236:1237, 1987). Promoter and enhancer elements have been isolated from a variety of eukaryotic sources including genes in yeast, insect, mammalian and plant cells. Promoter and enhancer elements have also been isolated from viruses and analogous control elements, such as promoters, are also found in prokaryotes. The
35 selection of a particular promoter and enhancer depends on the cell type used to express the protein of interest. Some eukaryotic promoters and enhancers have a broad host range while others are functional in a limited subset of cell types (*for review, see* Voss, *et al.*, Trends Biochem. Sci., 11:287, 1986; and Maniatis, *et al.*, *supra* 1987).

The terms "promoter element," "promoter" or "promoter sequence" refer to a DNA sequence

that is located at the 5' end (*i.e.* precedes) of the coding region of a DNA polymer. The location of most promoters known in nature precedes the transcribed region. The promoter functions as a switch, activating the expression of a gene. If the gene is activated, it is said to be transcribed, or participating in transcription. Transcription involves the synthesis of mRNA from the gene. The 5 promoter, therefore, serves as a transcriptional regulatory element and also provides a site for initiation of transcription of the gene into mRNA.

Promoters may be tissue specific or cell specific. The term "tissue specific" as it applies to a promoter refers to a promoter that is capable of directing selective expression of a nucleotide sequence of interest to a specific type of tissue (*e.g.*, petals) in the relative absence of expression of the same nucleotide sequence of interest in a different type of tissue (*e.g.*, roots). Tissue specificity of a promoter may be evaluated by, for example, operably linking a reporter gene to the promoter sequence to generate a reporter construct, introducing the reporter construct into the genome of a plant such that the reporter construct is integrated into every tissue of the resulting transgenic plant, and detecting the expression of the reporter gene (*e.g.*, detecting mRNA, protein, or the activity of a protein encoded by the reporter gene) in different tissues of the transgenic plant. The detection of a greater level of expression of the reporter gene in one or more tissues relative to the level of expression of the reporter gene in other tissues shows that the promoter is specific for the tissues in which greater levels of expression are detected. The term "cell type specific" as applied to a promoter refers to a promoter which is capable of directing selective expression of a nucleotide sequence of interest in a specific type of cell in the relative absence of expression of the same nucleotide sequence of interest in a different type of cell within the same tissue. The term "cell type specific" when applied to a promoter also means a promoter capable of promoting selective expression of a nucleotide sequence of interest in a region within a single tissue. Cell type specificity of a promoter may be assessed using methods well known in the art, *e.g.*, immunohistochemical staining. Briefly, tissue sections are embedded in paraffin, and paraffin sections are reacted with a primary antibody which is specific for the polypeptide product encoded by the nucleotide sequence of interest whose expression is controlled by the promoter. A labeled (*e.g.*, peroxidase conjugated) secondary antibody which is specific for the primary antibody is allowed to bind to the sectioned tissue and specific binding detected (*e.g.*, with avidin/biotin) by microscopy.

Promoters may be constitutive or regulatable. The term "constitutive" when made in reference to a promoter means that the promoter is capable of directing transcription of an operably linked nucleic acid sequence in the absence of a stimulus (*e.g.*, heat shock, chemicals, light, *etc.*). Typically, constitutive promoters are capable of directing expression of a transgene in substantially any cell and any tissue. In contrast, a "regulatable" promoter is one which is capable of directing a level of transcription of an operably linked nucleic acid sequence in the presence of a stimulus (*e.g.*, heat shock, chemicals, light, *etc.*) which is different from the level of transcription of the operably linked nucleic acid sequence in the absence of the stimulus.

The terms "infecting" and "infection" with a bacterium refer to co-incubation of a target biological sample, (*e.g.*, cell, tissue, *etc.*) with the bacterium under conditions such that nucleic acid

sequences contained within the bacterium are introduced into one or more cells of the target biological sample.

The term "*Agrobacterium*" refers to a soil-borne, Gram-negative, rod-shaped phytopathogenic bacterium which causes crown gall. The term "*Agrobacterium*" includes, but is not limited to, the strains *Agrobacterium tumefaciens*, (which typically causes crown gall in infected plants), and *Agrobacterium rhizogens* (which causes hairy root disease in infected host plants).
5 Infection of a plant cell with *Agrobacterium* generally results in the production of opines (e.g., nopaline, agropine, octopine etc.) by the infected cell. Thus, *Agrobacterium* strains which cause production of nopaline (e.g., strain LBA4301, C58, A208) are referred to as "nopaline-type"
10 *Agrobacteria*; *Agrobacterium* strains which cause production of octopine (e.g., strain LBA4404, Ach5, B6) are referred to as "octopine-type" *Agrobacteria*; and *Agrobacterium* strains which cause production of agropine (e.g., strain EHA105, EHA101, A281) are referred to as "agropine-type"
15 *Agrobacteria*.

The term "regulatory region" refers to a gene's 5' transcribed but untranslated regions, located immediately downstream from the promoter and ending just prior to the translational start of the gene.
15

The term "promoter region" refers to the region immediately upstream of the coding region of a DNA polymer, and is typically between about 500 bp and 4 kb in length, and is preferably about 1 to 1.5 kb in length.

20 In contrast, an "inducible" promoter is one which is capable of directing a level of transcription of an operably linked nucleic acid sequence in the presence of a stimulus (e.g., heat shock, chemicals, light, etc.) which is different from the level of transcription of the operably linked nucleic acid sequence in the absence of the stimulus.

25 The term "regulatory element" refers to a genetic element that controls some aspect of the expression of nucleic acid sequence(s). For example, a promoter is a regulatory element that facilitates the initiation of transcription of an operably linked coding region. Other regulatory elements are splicing signals, polyadenylation signals, termination signals, etc.

30 The enhancer and/or promoter may be "endogenous" or "exogenous" or "heterologous." An "endogenous" enhancer or promoter is one that is naturally linked with a given gene in the genome.
35 An "exogenous" or "heterologous" enhancer or promoter is one that is placed in juxtaposition to a gene by means of genetic manipulation (i.e., molecular biological techniques) such that transcription of the gene is directed by the linked enhancer or promoter. For example, an endogenous promoter in operable combination with a first gene can be isolated, removed, and placed in operable combination with a second gene, thereby making it a "heterologous promoter" in operable combination with the second gene. A variety of such combinations are contemplated (e.g., the first and second genes can be from the same species, or from different species).

The term "naturally linked" or "naturally located" when used in reference to the relative positions of nucleic acid sequences means that the nucleic acid sequences exist in nature in the relative positions.

The presence of "splicing signals" on an expression vector often results in higher levels of expression of the recombinant transcript in eukaryotic host cells. Splicing signals mediate the removal of introns from the primary RNA transcript and consist of a splice donor and acceptor site (Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor 5 Laboratory Press, New York [1989] pp. 16.7-16.8). A commonly used splice donor and acceptor site is the splice junction from the 16S RNA of SV40.

Efficient expression of recombinant DNA sequences in eukaryotic cells requires expression of signals directing the efficient termination and polyadenylation of the resulting transcript. Transcription termination signals are generally found downstream of the polyadenylation signal and 10 are a few hundred nucleotides in length. The term "poly(A) site" or "poly(A) sequence" as used herein denotes a DNA sequence which directs both the termination and polyadenylation of the nascent RNA transcript. Efficient polyadenylation of the recombinant transcript is desirable, as transcripts lacking a poly(A) tail are unstable and are rapidly degraded. The poly(A) signal utilized in an expression vector may be "heterologous" or "endogenous." An endogenous poly(A) signal is 15 one that is found naturally at the 3' end of the coding region of a given gene in the genome. A heterologous poly(A) signal is one which has been isolated from one gene and positioned 3' to another gene. A commonly used heterologous poly(A) signal is the SV40 poly(A) signal. The SV40 poly(A) signal is contained on a 237 bp *Bam*H/*Bcl*II restriction fragment and directs both termination and polyadenylation (Sambrook, *supra*, at 16.6-16.7).

20 The term "vector" refers to any genetic element, such as a plasmid, phage, transposon, cosmid, chromosome, retrovirus, virion, or similar genetic element, which is capable of replication when associated with the proper control elements and which can transfer gene sequences into cells and/or between cells. Thus, this term includes cloning and expression vehicles, as well as viral vectors.

25 The term "expression vector" as used herein refers to a recombinant DNA molecule containing a desired coding sequence (or coding sequences) - such as the coding sequence(s) for the hybrid enzyme(s) described in more detail below - and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host cell or organism. Nucleic acid sequences necessary for expression in prokaryotes usually include a promoter, an 30 operator (optional), and a ribosome binding site, often along with other sequences. Eukaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals. It is not intended that the present invention be limited to particular expression vectors or expression vectors with particular elements.

35 The term "transfection" refers to the introduction of foreign DNA into cells. Transfection may be accomplished by a variety of means known to the art including calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, glass beads, electroporation, microinjection, liposome fusion, lipofection, protoplast fusion, viral infection, biolistics (*i.e.*, particle bombardment) and the like.

The term "stable transfection" or "stably transfected" refers to the introduction and

integration of foreign DNA into the genome of the transfected cell. The term "stable transfectant" refers to a cell that has stably integrated foreign DNA into the genomic DNA.

The term "transient transfection" or "transiently transfected" refers to the introduction of foreign DNA into a cell where the foreign DNA fails to integrate into the genome of the transfected cell. The foreign DNA persists in the nucleus of the transfected cell for several days. During this time the foreign DNA is subject to the regulatory controls that govern the expression of endogenous genes in the chromosomes. The term "transient transfectant" refers to cells that have taken up foreign DNA but have failed to integrate this DNA.

The term "calcium phosphate co-precipitation" refers to a technique for the introduction of nucleic acids into a cell. The uptake of nucleic acids by cells is enhanced when the nucleic acid is presented as a calcium phosphate-nucleic acid co-precipitate. The original technique of Graham and van der Eb (Graham and van der Eb, *Virol.*, 52:456, 1973), has been modified by several groups to optimize conditions for particular types of cells. The art is well aware of these numerous modifications.

The terms "bombarding," "bombardment," and "biolistic bombardment" refer to the process of accelerating particles towards a target biological sample (*e.g.*, cell, tissue, *etc.*) to effect wounding of the cell membrane of a cell in the target biological sample and/or entry of the particles into the target biological sample. Methods for biolistic bombardment are known in the art (*e.g.*, U.S. Patent No. 5,584,807, the contents of which are incorporated herein by reference), and are commercially available (*e.g.*, the helium gas-driven microprojectile accelerator (PDS-1000/He, BioRad).

The term "microwounding" when made in reference to plant tissue refers to the introduction of microscopic wounds in that tissue. Microwounding may be achieved by, for example, particle bombardment as described herein.

The term "plant" as used herein refers to a plurality of plant cells which are largely differentiated into a structure that is present at any stage of a plant's development. Such structures include, but are not limited to, a fruit, shoot, stem, leaf, flower petal, *etc.* The term "plant tissue" includes differentiated and undifferentiated tissues of plants including, but not limited to, roots, shoots, leaves, pollen, seeds, tumor tissue and various types of cells in culture (*e.g.*, single cells, protoplasts, embryos, callus, protocorm-like bodies, *etc.*). Plant tissue may be *in planta*, in organ culture, tissue culture, or cell culture. Similarly, "plant cell(s)" may be cells in culture or may be part of a plant.

The term "transgenic" when used in reference to a cell refers to a cell which contains a transgene, or whose genome has been altered by the introduction of a transgene. The term "transgenic" when used in reference to a cell, tissue or to a plant refers to a cell, tissue or plant, respectively, which comprises a transgene, where one or more cells of the tissue contain a transgene (such as a gene encoding the hybrid enzyme(s) of the present invention), or a plant whose genome has been altered by the introduction of a transgene. Transgenic cells, tissues and plants may be produced by several methods including the introduction of a "transgene" comprising nucleic acid (usually DNA) into a target cell or integration of the transgene into a chromosome of a target cell by

way of human intervention, such as by the methods described herein.

The term "transgene" as used herein refers to any nucleic acid sequence which is introduced into the genome of a cell by experimental manipulations. A transgene may be an "endogenous DNA sequence," or a "heterologous DNA sequence" (i.e., "foreign DNA"). The term "endogenous DNA sequence" refers to a nucleotide sequence which is naturally found in the cell into which it is introduced so long as it does not contain some modification (e.g., a point mutation, the presence of a selectable marker gene, or other like modifications) relative to the naturally-occurring sequence. The term "heterologous DNA sequence" refers to a nucleotide sequence which is ligated to, or is manipulated to become ligated to, a nucleic acid sequence to which it is not ligated in nature, or to which it is ligated at a different location in nature. Heterologous DNA is not endogenous to the cell into which it is introduced, but has been obtained from another cell. Heterologous DNA also includes an endogenous DNA sequence which contains some modification. Generally, although not necessarily, heterologous DNA encodes RNA and proteins that are not normally produced by the cell into which it is expressed. Examples of heterologous DNA include reporter genes, transcriptional and translational regulatory sequences, selectable marker proteins (e.g., proteins which confer drug resistance), or other similar elements.

The term "foreign gene" refers to any nucleic acid (e.g., gene sequence) which is introduced into the genome of a cell by experimental manipulations and may include gene sequences found in that cell so long as the introduced gene contains some modification (e.g., a point mutation, the presence of a selectable marker gene, or other like modifications) relative to the naturally-occurring gene.

The term "transformation" as used herein refers to the introduction of a transgene into a cell. Transformation of a cell may be stable or transient. The term "transient transformation" or "transiently transformed" refers to the introduction of one or more transgenes into a cell in the absence of integration of the transgene into the host cell's genome. Transient transformation may be detected by, for example, enzyme-linked immunosorbent assay (ELISA) which detects the presence of a polypeptide encoded by one or more of the transgenes. Alternatively, transient transformation may be detected by detecting the activity of the protein (e.g., β -glucuronidase) encoded by the transgene (e.g., the *uid A* gene) as demonstrated herein (e.g., histochemical assay of GUS enzyme activity by staining with X-gluc which gives a blue precipitate in the presence of the GUS enzyme; and a chemiluminescent assay of GUS enzyme activity using the GUS-Light kit (Tropix)). The term "transient transformant" refers to a cell which has transiently incorporated one or more transgenes. In contrast, the term "stable transformation" or "stably transformed" refers to the introduction and integration of one or more transgenes into the genome of a cell. Stable transformation of a cell may be detected by Southern blot hybridization of genomic DNA of the cell with nucleic acid sequences which are capable of binding to one or more of the transgenes. Alternatively, stable transformation of a cell may also be detected by the polymerase chain reaction of genomic DNA of the cell to amplify transgene sequences. The term "stable transformant" refers to a cell which has stably integrated one or more transgenes into the genomic DNA. Thus, a stable transformant is

distinguished from a transient transformant in that, whereas genomic DNA from the stable transformant contains one or more transgenes, genomic DNA from the transient transformant does not contain a transgene.

The term "host cell" refers to any cell capable of replicating and/or transcribing and/or translating a heterologous gene. Thus, a "host cell" refers to any eukaryotic or prokaryotic cell (e.g., bacterial cells such as *E. coli*, yeast cells, mammalian cells, avian cells, amphibian cells, plant cells, fish cells, and insect cells), whether located *in vitro* or *in vivo*. For example, host cells may be located in a transgenic animal.

The terms "transformants" or "transformed cells" include the primary transformed cell and cultures derived from that cell without regard to the number of transfers. All progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. Mutant progeny that have the same functionality as screened for in the originally transformed cell are included in the definition of transformants.

The term "selectable marker" refers to a gene which encodes an enzyme having an activity that confers resistance to an antibiotic or drug upon the cell in which the selectable marker is expressed, or which confers expression of a trait which can be detected (e.g., luminescence or fluorescence). Selectable markers may be "positive" or "negative." Examples of positive selectable markers include the neomycin phosphotrasferase (NPTII) gene which confers resistance to G418 and to kanamycin, and the bacterial hygromycin phosphotransferase gene (*hyg*), which confers resistance to the antibiotic hygromycin. Negative selectable markers encode an enzymatic activity whose expression is cytotoxic to the cell when grown in an appropriate selective medium. For example, the HSV-*tk* gene is commonly used as a negative selectable marker. Expression of the HSV-*tk* gene in cells grown in the presence of gancyclovir or acyclovir is cytotoxic; thus, growth of cells in selective medium containing gancyclovir or acyclovir selects against cells capable of expressing a functional HSV TK enzyme.

The term "reporter gene" refers to a gene encoding a protein that may be assayed. Examples of reporter genes include, but are not limited to, luciferase (See, e.g., deWet *et al.*, *Mol. Cell. Biol.* 7:725, 1987 and U.S. Pat Nos., 6,074,859; 5,976,796; 5,674,713; and 5,618,682; all of which are incorporated herein by reference), green fluorescent protein (e.g., GenBank Accession Number U43284; a number of GFP variants are commercially available from CLONTECH Laboratories, Palo Alto, CA), chloramphenicol acetyltransferase, β -galactosidase, alkaline phosphatase, and horse radish peroxidase.

The term "overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. The term "cosuppression" refers to the expression of a foreign gene which has substantial homology to an endogenous gene resulting in the suppression of expression of both the foreign and the endogenous gene. As used herein, the term "altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

The terms "Southern blot analysis" and "Southern blot" and "Southern" refer to the analysis of DNA on agarose or acrylamide gels in which DNA is separated or fragmented according to size followed by transfer of the DNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized DNA is then exposed to a labeled probe to detect DNA species complementary to the probe used. The DNA may be cleaved with restriction enzymes prior to electrophoresis. Following electrophoresis, the DNA may be partially depurinated and denatured prior to or during transfer to the solid support. Southern blots are a standard tool of molecular biologists (J. Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, NY, pp 9.31-9.58, 1989).

The term "Northern blot analysis" and "Northern blot" and "Northern" refer to the analysis of RNA by electrophoresis of RNA on agarose gels to fractionate the RNA according to size followed by transfer of the RNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized RNA is then probed with a labeled probe to detect RNA species complementary to the probe used. Northern blots are a standard tool of molecular biologists (J. Sambrook, *et al.*, *supra*, pp 7.39-7.52, 1989).

The terms "Western blot analysis" and "Western blot" and "Western" refers to the analysis of protein(s) (or polypeptides) immobilized onto a support such as nitrocellulose or a membrane. A mixture comprising at least one protein is first separated on an acrylamide gel, and the separated proteins are then transferred from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized proteins are exposed to at least one antibody with reactivity against at least one antigen of interest. The bound antibodies may be detected by various methods, including the use of radiolabeled antibodies.

The term "antigenic determinant" refers to that portion of an antigen that makes contact with a particular antibody (*i.e.*, an epitope). When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies that bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as antigenic determinants. An antigenic determinant may compete with the intact antigen (*i.e.*, the "immunogen" used to elicit the immune response) for binding to an antibody.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated nucleic acid sequence" refers to a nucleic acid sequence that is identified and separated from one or more other components (*e.g.*, separated from a cell containing the nucleic acid, or separated from at least one contaminant nucleic acid, or separated from one or more proteins, one or more lipids) with which it is ordinarily associated in its natural source. Isolated nucleic acid is nucleic acid present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids are nucleic acids such as DNA and RNA which are found in the state they exist in nature. For example, a given DNA sequence (*e.g.*, a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs which encode a multitude of proteins. However, an isolated nucleic acid sequence comprising, for example, SEQ ID NO:1 includes, by way

of example, such nucleic acid sequences in cells which ordinarily contain, for example, SEQ ID NO:1 where the nucleic acid sequence is in a chromosomal or extrachromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid sequence may be present in single-stranded or double-stranded
5 form. When an isolated nucleic acid sequence is to be utilized to express a protein, the nucleic acid sequence will contain at a minimum at least a portion of the sense or coding strand (*i.e.*, the nucleic acid sequence may be single-stranded). Alternatively, it may contain both the sense and anti-sense strands (*i.e.*, the nucleic acid sequence may be double-stranded).

The term "purified" refers to molecules, either nucleic or amino acid sequences, that are
10 removed from their natural environment (or components of their natural environment), isolated or separated. An "isolated nucleic acid sequence" may therefore be a purified nucleic acid sequence. "Substantially purified" molecules are at least 60 % free, preferably at least 75 % free, and more
15 preferably at least 90 % free from other components with which they are naturally associated. As used herein, the term "purified" or "to purify" also refer to the removal of contaminants from a sample. The removal of contaminating proteins results in an increase in the percent of polypeptide of interest in the sample. In another example, recombinant polypeptides are expressed in plant,
20 bacterial, yeast, or mammalian host cells and the polypeptides are purified by the removal of host cell proteins; the percent of recombinant polypeptides is thereby increased in the sample. The present invention contemplates both purified (including substantially purified) and unpurified hybrid enzyme(s).

The term "composition comprising" a given polynucleotide sequence or polypeptide refers broadly to any composition containing the given polynucleotide sequence or polypeptide. The composition may comprise an aqueous solution. Compositions comprising polynucleotide sequences encoding GnTIII or fragments thereof may be employed as hybridization probes. In this case, the
25 GnTIII encoding polynucleotide sequences are typically employed in an aqueous solution containing salts (*e.g.*, NaCl), detergents (*e.g.*, SDS), and other components (*e.g.*, Denhardt's solution, dry milk, salmon sperm DNA, etc.).

The term "test compound" refers to any chemical entity, pharmaceutical, drug, and the like
30 that can be used to treat or prevent a disease, illness, sickness, or disorder of bodily function, or otherwise alter the physiological or cellular status of a sample. Test compounds comprise both known and potential therapeutic compounds. A test compound can be determined to be therapeutic by screening using the screening methods of the present invention. A "known therapeutic compound" refers to a therapeutic compound that has been shown (*e.g.*, through animal trials or prior experience with administration to humans) to be effective in such treatment or prevention.

35 As used herein, the term "response," when used in reference to an assay, refers to the generation of a detectable signal (*e.g.*, accumulation of reporter protein, increase in ion concentration, accumulation of a detectable chemical product).

The term "sample" is used in its broadest sense. In one sense it can refer to a animal cell or tissue. In another sense, it is meant to include a specimen or culture obtained from any source, as

well as biological and environmental samples. Biological samples may be obtained from plants or animals (including humans) and encompass fluids, solids, tissues, and gases. Environmental samples include environmental material such as surface matter, soil, water, and industrial samples. These examples are not to be construed as limiting the sample types applicable to the present invention.

5 The term "fusion protein" refers to a protein wherein at least one part or portion is from a first protein and another part or portion is from a second protein. The term "hybrid enzyme" refers to a fusion protein which is a functional enzyme, wherein at least one part or portion is from a first species and another part or portion is from a second species. Preferred hybrid enzymes of the present invention are functional glycosyltransferases (or portions thereof) wherein at least one part or 10 portion is from a plant and another part or portion is from a mammal (such as human).

15 The term "introduction into a cell" in the context of nucleic acid (*e.g.*, vectors) is intended to include what the art calls "transformation" or "transfection" or "transduction." Transformation of a cell may be stable or transient – and the present invention contemplates introduction of vectors under conditions where, on the one hand, there is stable expression, and on the other hand, where there is only transient expression. The term "transient transformation" or "transiently transformed" refers to the introduction of one or more transgenes into a cell in the absence of integration of the transgene into the host cell's genome. Transient transformation may be detected by, for example, enzyme-linked immunosorbent assay (ELISA) which detects the presence of a polypeptide encoded by one or more of the transgenes. Alternatively, transient transformation may be detected by detecting the 20 activity of the protein (*e.g.*, antigen binding of an antibody) encoded by the transgene (*e.g.*, the antibody gene). The term "transient transformant" refers to a cell which has transiently incorporated one or more transgenes. In contrast, the term "stable transformation" or "stably transformed" refers to the introduction and integration of one or more transgenes into the genome of a cell. Stable transformation of a cell may be detected by Southern blot hybridization of genomic DNA of the cell 25 with nucleic acid sequences which are capable of binding to one or more of the transgenes. Alternatively, stable transformation of a cell may also be detected by the polymerase chain reaction (PCR) of genomic DNA of the cell to amplify transgene sequences. The term "stable transformant" refers to a cell which has stably integrated one or more transgenes into the genomic DNA. Thus, a stable transformant is distinguished from a transient transformant in that, whereas genomic DNA 30 from the stable transformant contains one or more transgenes, genomic DNA from the transient transformant does not contain a transgene.

35 "Bisected oligosaccharide" shall be defined as an oligosaccharide comprising, *e.g.*, two mannose groups and another saccharide moiety attached to a mannose residue of the oligosaccharide. Examples of bisected oligonucleotides are given in Table 1.

DETAILED DESCRIPTION OF THE INVENTION

The GnTIII (for example, SEQ ID NO: 1, Figure 4A) expressed in the plant host cell of the present invention is a mammalian GnTIII. In a specific embodiment, the GnTIII is a human GnTIII (for example, SEQ ID NO: 2, Figure 4B). The GnTIII may also in a specific embodiment have 80 %

identity with the amino acid sequence of a mammalian GnTIII, more preferably at least about 90 %, even more preferably at least about 95 %, and most preferably at least about 97 % (hereinafter "homologous polypeptides"), which qualitatively retain the activity of said mammalian GnTIII. A polypeptide that has an amino acid sequence at least, for example, 95 % "identical" to a query amino acid sequence is identical to the query sequence except that the subject polypeptide sequence may include on average, up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95 % identical to a query amino acid sequence, up to 5 % of the amino acid residues in the subject sequence may be inserted, deleted or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the referenced sequence or in one or more contiguous groups within the reference sequence.

A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag, *et al.* (*Com. App. Biosci.* 6:237-245, 1990). In a sequence alignment, the query and subject sequence are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score= 1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

The GnTIII expressed in the plant host system of the present invention is encoded by a nucleic acid sequence that has at least 80 % identity with the nucleic acid sequence encoding an

amino acid sequence of a mammalian GnTIII, more preferably at least about 90 %, even more preferably at least about 95 %, and most preferably at least about 97 % (hereinafter "homologous polypeptides"), which qualitative retain the activity of said mammalian GnTIII. The nucleic acid sequence may be an RNA or DNA sequence.

5 A polynucleotide having 95 % "identity" to a reference nucleotide sequence of the present invention, is identical to the reference sequence except that the polynucleotide sequence may include, on average, up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95 % identical to a reference nucleotide sequence, up to 5 % of the nucleotides in the reference
10 sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5 % of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence, the ORF (open reading frame), or any fragment specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least
15 90 %, 95 %, 96 %, 97 %, 98 % or 99 % identical to a nucleotide sequence of the presence invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag, *et al.*, (*Comp. App. Biosci.*, 6:237-
20 245, 1990). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's (uridine) to T's (thymines). The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch
Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5,
25 Gap Size Penalty=0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

The invention also encompasses polynucleotides that hybridize to the nucleic acid sequence encoding said mammalian GnTIII. A polynucleotide "hybridizes" to another polynucleotide, when a single-stranded form of the polynucleotide can anneal to the other polynucleotide under the
30 appropriate conditions of temperature and solution ionic strength (see, Sambrook, *et al.*, *supra*). The conditions of temperature and ionic strength determine the "stringency" of the hybridization. For preliminary screening for homologous nucleic acids, low stringency hybridization conditions, corresponding to a temperature of 42 °C, can be used, *e.g.*, 5X SSC, 0.1 % SDS, 0.25 % milk, and no formamide; or 40 % formamide, 5X SSC, 0.5 % SDS). Moderate stringency hybridization
35 conditions correspond to a higher temperature of 55 °C, *e.g.*, 40 % formamide, with 5X or 6X SCC. High stringency hybridization conditions correspond to the highest temperature of 65 °C, *e.g.*, 50 % formamide, 5X or 6X SCC. Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the

nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of T_m (melting temperature) for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher T_m) of nucleic acid hybridizations decreases in the following order:

- 5 RNA:RNA, DNA:RNA, DNA:DNA.

Expression of GnTIII and other Heterologous Proteins in Plant Host Systems

In one embodiment, the nucleic acid encoding the mammalian GnTIII or other heterologous proteins, such as a heterologous glycoprotein or mammalian glycosyltransferase may be inserted into 10 an appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence, or in the case of an RNA viral vector, the necessary elements for replication and translation, as well as selectable markers. These include 15 but are not limited to a promoter region, a signal sequence, 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and transcription and translation termination sequences. Methods for obtaining such vectors are known in the art (see WO 01/29242 for review).

Promoter sequences suitable for expression in plants are described in the art, *e.g.*, WO 91/198696. These include non-constitutive promoters or constitutive promoters, such as, the 20 nopaline synthetase and octopine synthetase promoters, cauliflower mosaic virus (CaMV) 19S and 35S promoters and the figwort mosaic virus (FMV) 35 promoter (U.S. Pat. No. 6,051,753). Promoters used may also be tissue specific promoters targeted for example to the endosperm, aleurone layer, embryo, pericarp, stem, leaves, kernels, tubers, roots, etc.

A signal sequence allows processing and translocation of a protein where appropriate. The 25 signal can be derived from plants or could be non-plant signal sequences. The signal peptides direct the nascent polypeptide to the endoplasmic reticulum, where the polypeptide subsequently undergoes post-translational modification. Signal peptides can routinely be identified by those of skill in the art. They typically have a tripartite structure, with positively charged amino acids at the N-terminal end, followed by a hydrophobic region and then the cleavage site within a region of reduced 30 hydrophobicity.

The transcription termination is routinely at the opposite end from the transcription initiation 35 regulatory region. It may be associated with the transcriptional initiation region or from a different gene and may be selected to enhance expression. An example is the NOS terminator from Agrobacterium Ti plasmid and the rice α -amylase terminator. Polyadenylation tails may also be added. Examples include but are not limited to Agrobacterium octopine synthetase signal, (Gielen, *et al.*, EMBO J. 3:835-846, 1984) or nopaline synthase of the same species (Depicker, *et al.*, Mol. Appl. Genet. 1:561-573, 1982).

Enhancers may be included to increase and/or maximize transcription of the heterologous protein. These include, but are not limited to peptide export signal sequence, codon usage, introns, polyadenylation, and transcription termination sites (see, WO 01 /29242).

Markers include herbicide tolerance and prokaryote selectable markers. Such markers include resistance toward antibiotics such as ampicillin, tetracycline, kanamycin, and spectinomycin. Specific examples include but are not limited to streptomycin phosphotransferase (*spt*) gene coding for streptomycin resistance, neomycin phosphotransferase (*nptII*) gene encoding kanamycin or 5 geneticin resistance, hygromycin phosphotransferase (*hpt*) gene encoding resistance to hygromycin.

The vectors constructed may be introduced into the plant host system using procedures known in the art (reviewed in WO 01/29242 and WO 01/31045). The vectors may be modified to intermediate plant transformation plasmids that contain a region of homology to an *Agrobacterium tumefaciens* vector, a T-DNA border region from *A. tumefaciens*. Alternatively, the vectors used in 10 the methods of the present invention may be *Agrobacterium* vectors. Methods for introducing the vectors include but are not limited to microinjection, velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface and electroporation. The vector may be introduced into a plant cell, tissue or organ. In a specific embodiment, once the presence of a heterologous gene is ascertained, a plant may be regenerated 15 using procedures known in the art.

Uses of Mammalian GnTIII

The expression of mammalian GnTIII leads to bisected N-glycans on glycoproteins. Bisected N-glycans are important for biological activity of some mammalian glycoproteins. In 20 particular, bisected monoclonal antibodies have enhanced ADCC (antibody-dependent cellular cytotoxicity). Introduction of bisected structures leads to new or optimized functionalities and increased bioavailability of glycoprotein, e.g., increasing the antennary type increases half-life because of reduced clearance by the kidney. Accordingly, the invention is directed to a plant host system comprising said heterologous glycoprotein having bisecting oligosaccharides, particularly 25 bisecting G1cNAc residues and methods for producing said glycoprotein.

Furthermore, expression of GnTIII in plants leads to drastic increase of terminal G1cNAc's compared to wildtype plants (plant N-glycans contain far less G1cNAc residues compared to mammalian N-glycans). More G1cNAc residues on N-glycans of plant glycoproteins or recombinant glycoprotein produced in plants resembles mammalian N-glycans of glycoproteins. The introduction 30 of bisected G1cNAc in plant N-glycans (and in plant-produced recombinant glycoproteins such as Mabs) due to GnTIII expression in plants seems to prevent the N-glycan from degradation by β -N-acetylhexosaminidases. More G1cNAc residues means more acceptor substrate for β (1,4)-galactosyltransferase (GaIT) adding terminal galactose. Co-expression of GnTIII and a functional protein such as a transporter or a (mammalian) enzyme or functional fragment thereof providing N- 35 glycan biosynthesis, such as a galactosyltransferase, such as GAIT, or crossing GnTIII plants with GaIT plants and vice versa, leads to increased galactosylation of glycoproteins produced in these plants. Accordingly, the invention is directed to a plant host system comprising said mammalian GnTIII and said functional protein; the plant host system may further comprise a heterologous glycoprotein with increased galactosylation relative to a heterologous glycoprotein produced in a

plant host system not comprising said mammalian GnTIII and said functional protein, methods for providing said plant host systems and methods for producing said glycoprotein.

Generating stably transformed plants which produce tailored glycoproteins with commercial interest can be established by inoculating plant cells or tissues with *Agrobacterium* strains containing 5 a vector which comprises both nucleotide sequences encoding GnTIII, optionally N-glycosylation modifying enzymes and nucleotide sequences encoding commercially interesting heterologous glycoproteins or by the procedures described above such as electroporation, microinjection, velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface and electroporation. Alternatively, stably transformed plants which 10 produce tailored glycoproteins with commercial interest can be generated by simultaneous inoculation (cotransformation) of two or more *Agrobacterium* strains each carrying a vector comprising either nucleotide sequences encoding GNTIII, optionally, other N-glycosylation modifying enzymes or nucleotide sequences encoding glycoproteins of commercial interest or direct cotransformation of plant cells or tissues with said vectors. Alternatively, stably transformed plants 15 which produce tailored glycoproteins with commercial interest can be generated by (multiple) crossing(s) of plants with modified N-glycosylation with plants which express nucleotide sequences encoding proteins of commercial interest. In all of these procedures, the vector may also comprise a nucleotide sequence which confers resistance against a selection agent.

In order to obtain satisfactory expression of the proteins involved in N-glycosylation, GnTIII 20 and of the glycoproteins or polypeptides of commercial interest, the nucleotide sequences may be adapted to the specific transcription and translation machinery of the host plant as known to people skilled in the art. For example, silent mutations in the coding regions may be introduced to improve codon usage and specific promoters may be used to drive expression of the said genes in the relevant 25 plant tissues. Promoters which are developmentally regulated or which can be induced at will, may be used to ensure expression at the appropriate time, for example, only after plant tissues have been harvested from the field and brought into controlled conditions. In all these cases, choice of expression cassettes of the glycosylation modifying proteins and of the glycoproteins of commercial interest should be such that they express in the same cells to allow desired post translational modifications to the said glycoprotein.

As described above, in a specific embodiment, a plant that can be used in the method of the 30 present invention is a tobacco plant, or at least a plant related to the genus *Nicotiana*, however, use for the invention of other relatively easy transformable plants, such as *Arabidopsis thaliana* or *Zea mays*, or plants related thereto, is also particularly provided. For the production of recombinant 35 glycoproteins, the use of duckweed offers specific advantages. The plants are in general small and reproduce asexually through vegetative budding. Nevertheless, most duckweed species have all the tissues and organs of much larger plants including roots, stems, flowers, seeds and fronds. Duckweed can be grown cheaply and very fast as a free floating plant on the surface of simple liquid solutions from which they can easily be harvested. They can also be grown on nutrient-rich waste water, producing valuable products while simultaneously cleaning wastewater for reuse. Particularly

relevant for pharmaceutical applications, duckweed can be grown indoors under contained and controlled conditions. Stably transformed Duckweed can for example be regenerated from tissues or cells after (co)-inoculating with Agrobacterium strains containing each a (binary) vector which comprises one or more nucleotide sequences of interest encoding N-glycosylation modifying enzymes and/or genes encoding commercially interesting heterologous glycoproteins. The duckweed plant may, for example, comprise the genus *Spirodella*, genus *Wolffia*, genus *Wolffiella*, or the genus *Lemna*, *Lemna minor*, *Lemna miniscula* and *Lemna gibba*. Also mosses such as *Physcomitrella patens* offer advantages in that it can be grown cheaply under contained conditions. In addition the haploid genome of *Physcomitrella patens* is relatively easy to manipulate.

Expression in tomato fruits also offers specific advantages. Tomatoes can be easily grown in greenhouses under contained and controlled conditions and tomato fruit biomass can be harvested continuously throughout the year in enormous quantities. The watery fraction containing the glycoproteins of interest can be readily separated from the rest of the tomato fruit which allows easier purification of the glycoprotein. Expression in storage organs of other crops including but not limited to the kernels of corn, the tubers of potato and the seeds of rape seed or sunflower are also attractive alternatives which provide huge biomass in organs for which harvesting and processing technology is in place. Expression in nectar offers specific advantages with respect to purity and homogeneity of the glycoprotein secreted in the nectar.

Alternatively, a plant comprising a heterologous glycoprotein is crossed with a plant according to the invention comprising GnTIII and optionally at least one functional mammalian protein, e.g., a transporter or an enzyme providing N-glycan biosynthesis that is normally not present in plants, harvesting progeny from said crossing and selecting a desired progeny plant expressing said heterologous glycoprotein and expressing GnTIII and optionally a functional (mammalian) enzyme involved in mammalian-like N-glycan biosynthesis that is normally not present in plants. This process is known as crosspollination. In a preferred embodiment, the invention provides a method according to the invention further comprising selecting a desired progeny plant expressing said recombinant protein comprising bisecting oligosaccharide, particularly galactose residues and/or increased galactosylation. Now that such a plant is provided, the invention also provides use of a transgenic plant to produce a desired glycoprotein or functional fragment thereof, in particular wherein said glycoprotein or functional fragment thereof comprises bisecting oligosaccharide and/or increased galactosylation.

The invention additionally provides a method for obtaining a desired glycoprotein or functional fragment thereof comprising cultivating a plant according to the invention until said plant has reached a harvestable stage, for example when sufficient biomass has grown to allow profitable harvesting, followed by harvesting said plant with established techniques known in the art and fractionating said plant with established techniques known in the art to obtain fractionated plant material and at least partly isolating said glycoprotein from said fractionated plant material. The presence of desired proteins may be screened using methods known in the art, preferably using screening assays where the biologically active site is detected in such a way as to produce a

detectable signal. This signal may be produced directly or indirectly. Examples of such assays include ELISA or a radioimmunoassay.

The introduction of bisected G1cNAc residues due to expression of GnTIII can also be used for the prevention of removal (degradation) of saccharides from N-glycan by "blocking" activity 5 glycosidases, *e.g.*, β -N-acetylhexosaminidases and preventing the addition of other saccharides (driven by "other" subsequent glycosyltransferase genes) to N-linked glycan, *e.g.*, fucosylation, xylosylation. By controlling localization (*e.g.*, by providing other subcellular targetting signals) and/or controlling expression levels (*e.g.*, varying levels in independent transgenic plants or using different promoter) glycoform composition could be modulated. Hence introduction of bisecting 10 G1cNAc residues in glycoproteins in plants including recombinant glycoproteins, inhibits incorporation α -1,3-fucose by blocking activity α 1,3fucosyltransferase, α -1,4-fucose by blocking α -1,4-fucosyltransferase, β -1,2-xylose by blocking β -1,2-xylosyltransferase, β -1,3-galactose by blocking β -1,3-galactosyltransferase and removal/degradation of saccharides added to the N-glycan especially terminal G1cNAc residues by blocking activity of β -N-acetylhexosaminidases and 15 terminal β -1,4-galactose (added by expression of β -1,4-galactosyltransferase as provided by patent application WO 01/31045) by blocking β -1,4galactosidase. Thus in this way, controlled expression of GnTIII and controlled introduction of bisecting G1cNAc residues can be used to steer glycoform composition and/or limit glycoform heterogeneity.

20 Modified GnTIII andGnTIII Hybrid proteins

The invention is further directed to an isolated hybrid protein comprising a catalytic portion of mammalian GnTIII and a transmembrane portion of a protein, said protein residing in endoplasmic reticulum or Golgi apparatus of a eukaryotic cell. The invention is also directed to a modified mammalian GnTIII, wherein the transmembrane domain is removed but comprising a 25 retention signal such as KDEL for retention of said GnTIII in the ER.

A nucleic acid sequence encoding a hybrid enzyme comprising a transmembrane portion of a first enzyme and a catalytic portion of a second enzyme may be obtained as follows. The sequence 30 encoding the transmembrane portion is removed from the second enzyme, leaving a nucleic acid sequence comprising a nucleic acid sequence encoding the C-terminal portion of the second enzyme, which encompasses the catalytic site. The sequence encoding the transmembrane portion of the first enzyme is isolated or obtained via PCR and ligated to the sequence encoding a sequence comprising 35 the C-terminal portion of the second enzyme.

A nucleic acid sequence encoding a protein, particularly enzymes such as galactosyltransferases, mannosidases and N-acetylglucosamine transferases that are retained in the 35 ER may be obtained by removing the sequence encoding the transmembrane fragment and substituting it for a methionine (initiation of translation) codon and by inserting between the last codon and the stop codon of galactosyltransferase the nucleic acid sequence encoding an ER retention signal such as the sequence encoding KDEL (amino acid residue sequence: lysine-aspartic acid-glutamic acid-leucine) (Rothman, 1987).

Besides controlling expression, relocalization of GnTIII activity may also be controlled by making a fusion of the gene sequence coding for the enzymatic part of GnTIII with a transmembrane domain of other glycosyltransferases or enzymes/proteins residing in the endoplasmic reticulum (ER) or Golgi apparatus membrane, or by adding so-called retention signal such as but not limited to 5 KDEL for retention in the ER. Such relocalization modulates the addition of specific saccharides to the N-linked glycan of glycoproteins including recombinant glycoprotein and the prevention of removal of these.

The exchange of transmembrane domain of GnTIII with that of, for example, GnTI (TmGnTI), mannosidase II (TmManII) xylosyltransferase (TmXyl) or α -1,3 fucosyltransferase 10 (TmFuc) but not limited to these, enables earlier expression of GnTIII and introduction of bisecting G1cNAc at positions 20 to 22 in Figure 2. This prevents the action of subsequent glycosyltransferases such as xylosyltransferase and fucosyltransferase to act on the substrate leading to glycoforms lacking Xyl and Fuc. Importantly, the additional of terminal galactose by the action of 15 β -1,4)-galactosyltransferase (GaIT) is not inhibited by the bisecting G1cNAc- Co-expression of GaIT (Bakker, *et al.*, "Galactose-extended glycans of antibodies produced by transgenic plants" *Proc. Nat. Acad. Sci. USA* 98:2899-2904, 2001) results in structures similar as indicated to the right of the arrows annotated with 20, 21 and 22 in Figure 2. Although devoid of immunogenic xylose and 20 fucose residues, these structures have only one arm processed to complex type glycans. To allow conversion of also the other arm, in addition to relocating GnTIII, also Mannosidase II (ManII) and GnTII are relocated in the Golgi to act earlier in the glycan processing sequence. This can be 25 established in several ways. For example, by exchanging their respective transmembrane domains by that of GnTI (TmGnTI), which results in relocation to position indicated 5 in Figure 2. Alternatively, both ManII and GnTII can be relocalized to the ER by removing the transmembrane Golgi targeting domain and supplying the remaining enzyme fragments with a C-terminal ER 30 retention signal (*e.g.*, the amino acid residues KDEL). A plant expressing GaIT (Bakker, *et al.*, "Galactose-extended glycans of antibodies produced by transgenic plants" *Proc. Nat. Acad. Sci. USA* 98:2899-2904, 2001) as well as the relocated versions of GnTIII (*e.g.*, TmXyl-GnTIII), ManII (*e.g.*, TmGnTI-ManII) and GnTII (*e.g.*, TmGnTI-GnTII) can then be crossed with plants expressing the recombinant glycoprotein of interest (Figure 3) or can be retransformed with the gene encoding the 35 glycoprotein of interest such as the genes encoding an antibody. This allows the production of recombinant glycoproteins having bisected glycans with terminal galactose residues which are devoid of xylose and fucose. Transformation procedures and crossing (co-pollination) procedures are described above.

In another embodiment, GnTIII with transmembrane domain of Mannosidase II(TmManII-GnTIII) or xylosyltransferase (Tmxyl-GnTIII) combined with TmXyl-Ga1T, TmGnTI-GnTII, TmGnTI-ManII. This combination could either be obtained by coexpression or by combining through cross-pollination of the genes involved and leads to glycoproteins including recombinant glycoproteins, lacking xylose and fucose on the core sequence but having bisected GIcNAc residues on the trimannosyl core and terminal galactose.

EXAMPLES

The effect of the introduction of GnTIII in plants on the occurrence of bisected oligosaccharides on the glycans of plant glycoproteins has been evaluated. The human gene for GnTIII has been cloned, and a C-terminal c-myc tag for analysis of expression of the tagged fusion protein has been provided and the whole has been placed under control of plant regulatory elements for introduction in tobacco. It is shown that GnTIII is expressed in plants and that expression results in bisected oligosaccharide structures on endogenous plant glycoproteins. The amount of N-glycans containing at least two G1cNAc residues more than doubled compared to those found in normal tobacco plants. Remarkably, the expression of GnTIII also resulted in a significant reduction of complex type N-glycan degradation products as apparent from matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) analyses of the isolated glycans of endogenous plant glycoproteins. These data suggest that expression of GnTIII in tobacco resulting in the introduction of bisected structures on N-glycans protects the glycans from degradation by β -N-acetylhexosaminidases. β -N-acetylhexosaminidases have broad specificity for non-reducing terminal G1cNAc and β -N-acetylglucosamine (GalNAc) cleaving amongst others G1cNAc- β 1-2 linkages typically present on the trimannosyl core (Man- α -1-3 and Man- α -1-6).

Example 1

Plasmids and plant transformation. PAC clone RP5-1104E15 GnTIII (SEQ ID NO: 1, Figure 4A) was obtained from Pieter J. de Jong, Children's Hospital Oakland Research Institute (CHORI) and is available on request through Sanger Center being part of clone set HBRC_1.sc. The clone originates from Homo sapiens, male, blood and can be requested through <http://www.sanger.ac.uk/Teams/Team63/CloneRequest/> (from Human chromosome 22q12.3-13.1; The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK; www.sanger.ac.uk)

The human gene for GnTIII was cloned from said PAC clone by PCR using AccuTaq LA DNA polymerase (SigmaAldrich) and primers GNT3F (5'-atactcgagttacaataatgaagatgagacgct-3'; SEQ ID NO: 3) and GNT3Rmyc (5'-tatggatcctaattcagatccttctgagatgag-3'; SEQ ID NO: 4). Oligos were from Eurogentec (Belgium). PCR was performed on a PerkinElmerCetus 480 thermal cycler (ABI/PE) using optimal conditions for the AccuTaq polymerase according to the manufacturer. The resulting fragment was cloned in EcoRV site of pBluescribe SK+ (Stratagene, Inc., La Jolla, CA USA) and sequence verified. Sequencing was performed using fluorescently labelled dideoxynucleotides essentially as described (Sanger, *et al.*, "DNA sequencing with the dideoxy chain-terminating inhibitors" *Proc. Nat. Acad. Sci. USA* 74:5463-5467, 1977) and reaction mixtures were run on an Applied Biosystems 370A or 380 automated DNA sequencer. Data were analysed using different software modules freely available on the web and compared with the DNA sequence of human GnTIII present in the database.

A 1.6 kb HpaI/BamHI fragment containing the GnTIII gene with C-terminal c-myc tag was

subsequently cloned into the Sma/BglII site of pUCAP35S (Van Engelen, *et al.*, "Coordinate expression of antibody subunit genes yields high levels of functional antibodies in roots of transgenic tobacco" *Plant Molecular Biology* 26:1701-1710, 1994) and named pAMV-GnTIII. The cauliflower mosaic virus 35S (CaMV35S) 20 promoter expression cassette with modified GnTIII gene was
5 subsequently cloned as a Ascl/PacI fragment in the binary vector pBINPLUS (Van Engelen, *et al.*, "Coordinate expression of antibody subunit genes yields high levels of functional antibodies in roots of transgenic tobacco" *Plant Molecular Biology* 26:1701-1710, 1994) resulting in pBINPLUSGnTIII and introduced in *Agrobacterium tumefaciens* strain Ag10 by electroporation. Transformation of
10 *Nicotiana tabacum* variety Samsun NN was as described before (Horsch, *et al.*, "A simple and general method for transferring genes into plants" *Science* 227:1229-1231, 1985). Sixteen independent transgenic plants were selected and grown to maturity in the greenhouse as described. Leaf material was analysed for expression of GnTIII (SEQ ID NO: 2, Figure 4B) and glycan composition of endogenous cellular glycoproteins.
15

Example 2

Analysis of expression. Total protein extracts of tobacco leaves were prepared as described before (Bakker, *et al.*, "Galactose-extended glycans of antibodies produced by transgenic plants" *Proc. Nat. Acad. Sci. USA* 98:2899-2904, 2001). The amount of protein present in samples was estimated by the Bradford method (Bradford, M.M., "A rapid and sensitive method for the
20 quantitation of microgram quantities of protein utilizing the principle of protein-dye binding" *Anal Biochem* 72:248-254, 1976) using bovine serum albumin as standard. Fixed amounts of protein samples were run on precast 10 or 12 % SDS-PAGE gels (Bio-Rad) under reduced conditions. Rainbow coloured molecular weight protein markers were from Amersham. Western blot analysis was performed essentially as described (Bakker, *et al.*, "Galactose-extended glycans of antibodies
25 produced by transgenic plants" *Proc. Nat. Acad. Sci. USA* 98:2899-2904, 2001). Separated proteins were transferred to nitrocellulose (BA85, Schleicher and Schuell or Trans-Blot Transfer Medium, Bio-Rad) using a Bio-Rad Mini Trans-blot Electrophoretic Transfer Cell in 3[cyclohexylamino]-1-propanesulfonic acid (CAPS) buffer for 60 min. Expression of the GnTIII-c-myc fusion protein was analysed by affinoblotting using a peroxidase labelled c-myc antibody. Introduction of bisecting
30 oligosaccharides in endogenous tobacco glycoproteins was visualized by incubation with biotinylated erythroagglutinating phytohemagglutinin (E-PHA; Vector Laboratories). Detection was performed by enhanced chemiluminescence using Lumi-Light Western Blotting Substrate from Roche (Roche Diagnostics GmbH, Mannheim, Germany) on a Lumi-Imager F 1 apparatus (Boehringer Mannheim GmbH, Mannheim, Germany) using LumiAnalyst software (version 3.0).

35

Example 3

Matrix-assisted Laser Desorption Ionization Time-of-Flight (MALDI-TOF) Mass Spectrometry. For the analysis of glycan structure cellular proteins were isolated from tobacco leaves of a selected plant transformed with human GnTIII (GnTIII-17). Protein isolation and N-

glycan preparation were performed as described (Elbers, *et.al.*, 2001). The N-glycans were desalted on a nonporous, graphitized carbon-black column (Carbograph Ultra Clean Tubes, Alltech Associates) before mass spectrometry analysis as described. MALDI-TOF spectra were measured on a Micromass (Manchester, U.K.) Tof spec E MALDI-TOF mass spectrometer. Mass spectra were 5 performed in positive mode by using 2,5-dihydroxybenzoic acid as the matrix essentially as described (Elbers, *et al.*, 2001).

Expression of human GnTIII introduces bisecting N-glycans on endogenous glycoproteins in *N. tabacum*. Human GnTIII was introduced in tobacco plants by *Agrobacterium*-mediated transformation of binary vector pBINPLUSGnTIII containing a cDNA harbouring the complete 10 coding sequence fused to a C-terminal c-myc tag under control of the constitutive CaMV35S promoter. Sixty independent transgenic shoots selected for kanamycin resistance were obtained which were analysed for expression of the GnTIII-c-myc fusion protein using the c-myc antibody. Analysis revealed that all expressed the gene at various levels. Fourteen were selected, rooted and transferred to the greenhouse. One plant (GnTIII-17) selected for high expression of the GnTIII-c- 15 myc fusion protein using the c-myc antibody was analysed for the occurrence of bisected GlcNAc residues on N-glycans of endogenous tobacco glycoproteins using a specific binding assay with the biotinylated lectin E-PHA. SDS-PAGE of protein extracts followed by transfer to nitrocellulose and analysis using the specific binding assay with the biotinylated E-PHA lectin revealed that 20 endogenous tobacco glycoproteins of GnTIII-17 contained bisected oligosaccharides whereas those of control tobacco did not. GnTIII-17 was multiplied in the greenhouse for further detailed analysis of glycan structure by MALDI-TOF.

Example 4

Oligosaccharide distributions and level of bisected complex oligosaccharides in 25 wildtype and selected transgenic GnTIII-17 tobacco plant. Endogenous glycoproteins were isolated from young leaves of control tobacco plant and the selected GnTIII-17 plant to investigate in detail the effect of expression of human GnTIII on the structure of glycans N-linked to glycoproteins. A comparison of the structures of the N-glycans isolated from glycoproteins present in leaves of 30 control wild-type tobacco plants with those from plant GnTIII-17 using MALDI-TOF is represented in Figure 1. MALDI-TOF allows for the detection of different molecular species in the pool of the N-glycans (glycoforms) and shows a mixture of ions that were assigned to (M+Na)+ adducts of high-mannose (Man)- type N-glycans ranging from **d**, Mans to **n**, Man9 and of mature N-glycans from the truncated structure **a**, XM3GN2 to **m**, GN3FXM3GN2 (for structure see Table 1; for a summary of 35 the data see, Table 2). In addition to the N-glycans characterized in the control plants (Figure 1A), the MALDI-TOF MS of the glycan mixture from plant GnTIII-17 (Figure 1B) showed at least two ions assigned to N-linked glycans that result from the action of the human GnTIII enzyme (for a comparison see Table 1 and Table 2). These oligosaccharides, GN3XM3GN2 (**i**) and GN3FXM3GN2 (**k**) representing 8 % and 31 % respectively of the population, contain three GlcNAc residues each linked to one of the three mannoses of the trimannosyl core structure of the

N-linked glycan.

Analysis of glycan structure through MALDI-TOF as performed here cannot distinguish between G1cNAc residues β (1,2)- or β (1,4)-linked to mannose. Hence, it was not clear if or to what extent the structures GN2XM3GN2 and GN2FXM3GN2 have bisecting oligosaccharides.

5 Additional experiments are required to reveal that these structures are a mix of normal and bisected oligosaccharides or a single compound.

In the light of the observed lethality of CHO cell that overexpress GnTIII (Umana, *et al.*, “Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity” *Nature Biotechnology* 17:176-180, 1999), remarkably transgenic plants having 10 significant amounts of bisected glycans look phenotypically normal and are completely fertile (can be cross-pollinated and self-pollinated).

Example 5

Expression of human GnTIII in tobacco seems to protect N-glycans from degradation by D-Nacetylhexosaminidases and more than doubles terminal N-glucosamination. MALDI-TOF analysis of extracts clearly showed that at least 40 % of the population of glycoforms now has a bisecting G1cNAc in complex-type N-linked glycans of cellular tobacco proteins through the action of the GnTIII enzyme. Moreover 70 % of the population of complex-type N-linked glycans of endogenous glycoproteins of GnTIII-17 has two or three terminal GIcNAc residues compared to 20 about 30 % for wildtype tobacco (Table 1). The observed *de novo* synthesis of at least 40 % bisected complex-type N-linked glycans upon expression of GnTIII in tobacco (Figure 1B, Table 1 and Table 2) coincides with the disappearance of mainly FXM3GN2 (b, from 30 % to 4 %) and GNFXM3GN2 (f, from 10 to 2%) and to a minor degree GN2FXM3GN2 (j, from 29 % to 19 %). In addition it also coincides with a significant increase in GN2XM3GN2 (h) from 4 % in wildtype tobacco to 14 % in 25 GnTIII-17. Whether the latter GN2XM3GN2 (h) in GnTIII plants has the second G1cNAc linked to the (β -linked mannose of the trimannosyl core of the N-linked glycan and hence is the result of GnTIII activity, or to the second α -linked mannose of the trimannosyl core remains to be investigated (see above).

Saccharides a, b and c accounting for 40 % of the N-linked glycans in wildtype tobacco 30 plants, are degradation products expected to have arisen from mature glycans of endogenous tobacco glycoproteins after GnTI activity since an *Arabidopsis thaliana* mutant lacking GnTI activity did not contain xylose and fucose residues in the N-glycans of endogenous glycoproteins (Von Schaewen, *et al.*, “Isolation of a mutant *Arabidopsis* plant that lacks N-Acetylglucosaminyl transferase I and is unable to synthesize golgi-modified complex N-linked glycans” *Plant Physiology* 102:1109-1118, 35 1993). The 7-fold decrease (40 % > 6 %) in these structures in GnTIII-17 together with the threefold reduction of GNXM3GN2 and XM3GN2 (12 % > 4 %) suggests that the introduction of a bisected GIcNAc protects the mature N-linked glycan from degradation by endogenous glycosidases especially β -Nacetylhexosaminidases that removes terminal G1cNAc. The total amount of N-linked glycans expected to have arisen from degradation of mature, full-length N-linked glycans has hence

decreased fivefold (from 52 % to 10 %).

Example 6

Vector construction and DNA preparation for maize transformation.

5 The human GNTIII gene along with its 3' c-myc immunodetection tag was obtained by PCR from plasmid pAMV-GNTIII by the following method. Primers MS20 and MS19 homologous to the 5' and 3' ends of the hGNTIII gene respectively, were designed and synthesized to add a Pme1 site and a stop codon to the 3' end of the gene.

10 MS20 (5' *Nco*1 site): 5'-CCATGGTGATGAGACGCTAC-3' (SEQ ID NO: 5)
MS19 (adds stop and *Pme*1 site 3'): 5'-GTTTAAACCTAGGATCCTAATTAGTCAGATCCTCT-3' (SEQ ID NO: 6)

15 Following gel electrophoresis to identify the correct sized PCR product, the 1.6 kbp PCR product was recovered from the gel with a QIAquick Gel Extraction Kit (Qiagen, Valencia, CA). Plasmid 4005 (see, SEQ ID NO: 8) (Figures 5A and 5B), which contains a Zmubi /GUS /per5 cassette (Christensen, *et al.*, *Plant Molec. Biol.* 18:675-689, 1992), was digested with Nco1 and Pme1 to release the GUS gene and the vector fragment was recovered from a gel with a QIAquick Gel Extraction Kit (Qiagen, Valencia, CA).

20 Following digestion with Nco1 and Pme1, the PCR-derived hGNTIII fragment was ligated to the vector fragment left after digestion of pDAB4005 with Nco1 and Pme1, to create the intermediate plasmid pDAB7119 (see, SEQ ID NO: 9) (Figures 6A and 6B). Intermediate plasmid pDAB7119 was cut with Spe1 and Sph1 to release the hGNTIII plant expression cassette, which was treated with T4 DNA polymerase to create blunt ends.

25 Plasmid pDAB8504 (SEQ ID NO: 10) (see, Figures 7A and 7B), which contains the RB7 MAR sequences, was digested with Srf1 and blunt ended with T4 DNA polymerase. Following treatment with calf intestinal phosphatase, the treated 8504 fragment and the hGNTIII plant expression cassette were ligated to create plasmid pDAB 7113 (SEQ ID NO: 10) (see, Figures 8A and 8B), which contains RB7 MAR sequences flanking the gene of interest and the selectable marker cassette as follows: RB7 MAR // Zmubi promoter/hGNTIII/per5 3'UTR // Rice actin promoter (D. McElroy, *et al.*, "Isolation of an efficient actin promoter for use in rice transformation" *The Plant Cell* 2:163-171, 1990) /PAT/ Zm lipase 3'UTR // RB7 MAR.

35 The integrity of the GNTIII sequence was checked by sequencing (Big Dye Terminator Cycle Sequencing Ready Reaction, Applied Biosystems, Foster City, CA) and was confirmed to encode the human GNTIII enzyme. One base change, G384 → 384, was found but this substitution does not affect the encoded amino acid, proline 128.

Plasmid pDAB7113 was grown up in 2 L of medium (LB + amp) and purified with Qiagen plasmid Giga kit to produce 5 milligrams of purified plasmid for plant cell transformation.

Example 7

Transformation of maize cells. Plasmid pDAB7113 was introduced into maize cells with WHISKERS-mediated DNA transfer essentially as described in these citations, and as follows (Frame, B., *et al.*, "Production of fertile transgenic maize plants by silicon carbide whisker-mediated transformation" *Plant J.* 6:941-948, 1994; Thompson, J., *et al.*, "Maize transformation utilizing silicon carbide whiskers: a review" *Euphytica* 85:75-80, 1995; P. Song, C. Q. Cai, M. Skokut, B. Kosegi, and J. Petolino, "Quantitative real-time PCR as a screening tool for estimating transgene copy number in Whiskers-derived transgenic maize" *Plant Cell Rep.* 20:948-954, 2002; both of which are incorporated herein by reference).

Embryogenic maize suspension cell cultures were subcultured on medium G-N6 (N6 medium containing 30 gm/L sucrose, 100 mg/L inositol, and 2 mg/L 2,4-D) the day before whisker mediated transformation. On the day of the experiment, cells were pretreated with osmoticum by shaking with medium G-N6 containing 0.2 Molar each mannitol and sorbitol for 30 minutes. Thirty six mls of cells were transferred to a 250 ml centrifuge bottle in 50 ml of medium G-N6, to which was added 8.1 ml of a 5 % (w/v) silicon carbide whiskers suspension (Silar SC-9, Advanced Composite Materials, Greer, S.C.) in medium, plus 170 ul of 1 mg/ml plasmid solution (in TE buffer). The centrifuge bottle containing cells, whiskers and DNA was agitated vigorously on a modified Red Devil brand paint mixer for 10 seconds. Whiskered cells were then shaken for two hours in medium with half the level of added osmoticum. Whiskered cells were recovered by filtration on a sterile Buchner funnel and the filter papers were placed on semisolid G-N6 medium for 1 week. After 1 week the filters were moved to semisolid G-N6 medium containing 1 mg/L Herbiace (a commercial formulation of 20% bialaphos, Meiji Seika, Tokyo, Japan). Two weeks later, the cells were removed from the filter paper, mixed with melted G-N6 + 1 mg/L Herbiace (G-N6 +1H) medium also containing 7 gm/L Seaplaque agarose (BioWhittaker, Rockland, Maine), and spread on top of G-N6 + 1H solid medium. Plates were cultured in the dark at 30 °C. Colonies resistant to the selective agent were recovered 5-7 weeks post embedding, and individually moved to fresh G-N6+1H medium for further increase of tissue mass.

Example 8

Molecular analysis for copy number of inserted DNA. Tissue from each transgenic isolate was individually freeze-dried in a lyophilizer and DNA was extracted by a standard method (DNAeasy 96 Plant Kit, Qiagen). The copy number of inserted transgenic DNA was estimated by the Invader Operating System, available from Third Wave Technologies (Third Wave Technologies, Madison, Wisconsin, twt.com). Primers were designed by the Third Wave Technologies company specifically for the PAT selectable marker and its copy number was estimated relative to genomic DNA copy number for the endogenous maize alpha-tubulin gene.

Example 9

Test transgenic maize callus for altered lectin binding due to expression of the GntIII

gene. Callus samples from 100 individually isolated unique transgenic events were extracted as follows. Samples from each event were fresh frozen in 96-well cluster tube boxes (Costar 1.2 ml polypropylene, with lid) along with a steel and a tungsten bead in each well. 450 ul of extraction buffer (25 mM sodium phosphate pH6.6, 100 mM NaCl, 30 mM sodium bisulfate, 1 % v/v Triton X-100) was added per well and the box of samples was pulverized for 3 minutes full speed on a Kleco Bead Mill. The plate was centrifuged (4 °C) at 2500 rpm for 10 minutes. Extracts were removed to a 96-well deep well plate and frozen for storage. All screening assays were performed on these extracts of individual events.

Protein analyses (microtiter plate protocol, BioRad 500-0006) were made to determine the total protein for each extract. 25 ug protein per sample were loaded in 20 ul loading buffer (Laemmli, U.K. Nature 277:680 (1970)). Gels (4-20 % Criterion PAGE gels, 12+2 wells, BioRad 345-0032) were electrophoresed at 65 mA in Tris/glycine /SDS running buffer (BioRad 161-077). After soaking in transfer buffer (running buffer plus 20 % v/v methanol) for 10 minutes, the gels were transferred to nitrocellulose membranes using a semi-dry blotter (150 mA/1.5 hrs). The membranes were incubated for 30 minutes in blocking buffer (20 mM Tris, 144 mM NaCl, 0.5 % v/v Tween 20, 10 % w/v nonfat powdered milk) at room temperature, then the blocking buffer was removed and replaced with the primary detection lectin (*Phaseolus hemagglutinin E*, biotinylated, Vector Laboratories B-1125) 2.5 ug/ml in blocking buffer. The primary detection lectin was incubated on the membrane for 1 hour at room temperature. The primary detection solution was removed, the membrane was rinsed once with blocking buffer and the secondary detection solution was added (avidin-HRP, BioRad 170-6528, at 1:5000, plus molecular weight marker detection agent, StrepTactin-HRP, BioRad 161-0380 at 1:10,000 in blocking buffer. The secondary detection reagent was incubated on the membrane for 1 hour at room temperature. During the blocking, primary, and secondary reagent steps the solutions were mixed on the blots. The secondary detection reagent was then removed and the membrane was rinsed with Tris buffered saline (20 mM Tris, 144 mM NaCl) containing 0.5 % Tween 20 three times at 10 minutes each and once more for 5 minutes. After dripping off the excess rinse solution, the blot was soaked in substrate ECL (Pierce 34080) for 1 minute, excess ECL solution was drained off, and the membrane was exposed to film. Negative controls were included in each gel to discriminate new glycoprotein bands now visible with this bisecting glycan -detecting lectin reagent on the transgenic callus extracts.

Positive test results (Table 5) for the E-PHA binding were rated as 0 (negative), 1 (one plus, weak) 2 (two pluses, moderately strong) or 3 (three pluses, strongest rating). Callus of events rated 2 or 3 were selected to produce sample for mass analysis. Samples 25, 26, 33, 48, 55, 56 and 59 were pooled to produce the protein extract for MALDI-TOF analysis of glycan substructures. A gel blot example (Figure 12) shows samples 19 through 27.

Example 10

Test transgenic maize callus for c-myc epitope expression. Callus samples from 100 individually isolated unique transgenic events were extracted as follows. Samples from each event

were fresh frozen in 96-well cluster tube boxes (Costar 1.2 ml polypropylene, with lid) along with a steel and a tungsten bead in each well. 450 ul of extraction buffer (25 mM sodium phosphate pH6.6, 100 mM NaCl, 30 mM sodium bisulfate, 1 % v/v Triton X-100) was added per well and the box of samples was pulverized for 3 minutes full speed on a Kleco Bead Mill. The plate was centrifuged (4 °C) at 2500 rpm for 10 minutes. Extracts were removed to a 96-well deep well plate and frozen for storage. All screening assays were performed on these extracts of individual events.

Protein analyses (microtiter plate protocol, BioRad 500-0006) were made to determine the total protein for each extract. 25 ug protein per sample were loaded in 20 ul loading buffer (Laemmli, U.K. Nature 277:680, 1970)). Gels (4-20 % Criterion PAGE gels, 12+2 wells per gel, BioRad 345-0032) were electrophoresed at 65 mA in Tris/glycine/SDS running buffer (BioRad 161-0772). After soaking in transfer buffer (running buffer plus 20 % methanol) for 10 minutes, the gels were transferred to nitrocellulose membranes using a semi-dry blotter (150 mA/1.5 hrs). The membranes were incubated for 30 minutes in blocking buffer (20 mM Tris, 144 mM NaCl 0.5 % v/v Tween 20, 10 % w/v dry milk) at room temperature, then the blocking buffer was removed and replaced with the primary detection reagent, Mouse anti-c-myc clone 9E10 (sigma M5546) at 1 ug/ml in blocking buffer. After 1 hour of incubation at room temperature, the primary detection reagent was removed and the membrane was rinsed with blocking buffer. The secondary detection reagent, anti-mouse - HRP (BioRad 170-6516) at 1:10,000 plus a molecular weight marker detection reagent (StrepTactin - HRP, BioRad 161-0380) at 1:10,000 in blocking buffer, was then added and incubated on the membrane for 1 hour at room temperature. During the blocking, primary, and secondary reagent steps the solutions were mixed on the blots. The secondary detection agent was removed, and the membrane was rinsed three times with Tris buffered saline (20 mM Tris, 144 mM NaCl) containing 0.5 % Tween 20 for 10 minutes each, plus another 5 minute rinse. After draining off the excess rinse solution the membrane was soaked in ECL reagent (Pierce 34080) for 1 minute, drained, and then exposed to film.

As detailed above, callus samples from independent events 1-100 were screened for expression of the c-myc epitope. Then, samples 3, 11, 12, 26, 31, 55 and 64 were analysed and showed the presence of a band in the predicted molecular weight range of 50-55 kilodaltons. These callus samples were pooled to produce a protein sample for glycan analysis by MALDI-TOF. A representative blot is shown in Figure 13.

Example 11

Preparation of extract for mass spec analysis of glycans. The samples were prepared from combined calluses of several maize callus events which tested positive for GnTIII transgene expression based on lectin blotting using E-PHA. Callus tissue was collected fresh and stored frozen at -80 °C, then ground to a fine powder in liquid nitrogen. Weighed sample was added to extraction buffer (5 mM EDTA, 0.5 mM PMSF, 20 mM sodium bisulfite, 150 mM sodium phosphate buffer pH 7.4, and 0.4 mM PVPP soluble MW 40,000) and stirred for 30 minutes at 4 °C. After centrifugation at 5000 x G at 4 °C, the supernatant was collected. Ammonium sulfate and wash buffer (5 mM

EDTA, 150 mM sodium phosphate buffer, pH 7.4) were added to the supernatant to achieve a final concentration of 20 % (w/v) ammonium sulfate. After centrifugation 5 minutes at 5000 x G at 4 °C, the supernatant was transferred to a fresh tube and additional ammonium sulfate plus wash buffer were added to achieve 60 % (w/v) ammonium sulfate. This preparation was stirred overnight at 4 °C, 5 then centrifuged 20 minutes at 10,000 x G . The pellet was recovered in 5 ml of wash buffer and frozen at -80 °C, then lyophilized at 4 °C until dry. Samples were sent to the lab for glycan analysis.

Example 12

Maize plant regeneration from transgenic callus tissue. For plant regeneration from 10 transformed callus, tissue was placed onto regeneration media containing MS basal salts and vitamins (Murashige T. and F. Skoog , *Physiol Plant* 15:473-497, 1962), 30 g/l sucrose, 5 mg/l 6-benzylaminopurine (BA), 0.025 mg/l 2, 4-dichlorophenoxyacetic acid (2,4-D), 1 mg/l Herbiace (a commercial formulation of 20 % bialaphos, Meiji Seika, Tokyo, Japan), and 2.5 g/l Gelrite, pH 5.7. 15 Cultures were grown in the light. When shoots reached 1-3 cm in length, they were transferred into vessels containing SH basal salts and vitamins (Schenk R. and A.C. Hildebrandt, *Can J Bot* 50:199-204, 1972), 10 g/l sucrose, 1 g/l myo-inositol, and 2.5 g/l Gelrite, pH 5.8.

Plants were screened for expression of GNTIII by altered binding of the lectin E-PHA to 20 endogenous proteins. Samples were then screened for E-PHA binding as described in Example 9, *supra*. The protein extract and 20 % / 60 % ammonium sulfate precipitate was prepared exactly as for 25 the callus samples as described in Example 13, *infra*. One plant each from plants regenerated from 23 independent events were screened by lectin blotting for the results of expression of the GNTIII gene. Four of these events gave positive signals for E-PHA binding. These four events had also tested positive at the callus stage. Plants regenerated from those four events were pooled to produce a protein extract for glycan analysis by MALDI-TOF.

25

Example 13

Oligosaccharide distributions and level of bisected complex oligosaccharides in wildtype 30 and selected transgenic corn calli. Endogenous glycoproteins were isolated from control corn calli and selected corn calli expressing GnTIII based on lectin blotting using E-PHA. In addition, the present invention also contemplates the extraction of c-myc tagged samples. E-PHA and c-myc 35 tagged samples may be callus, plant cells, plant tissues or entire plants as defined in the definitions section *supra*. A comparison of the structures of the N-glycans isolated from glycoproteins present in calli is presented in Figures 9A and 9B. MALDI-TOF allowed for the detection of different molecular species in the pool of the N-glycans (glycoforms) and showed a mixture of ions that were assigned to (M+Na)+ adducts of high-mannose (Man)- type N-glycans ranging from d, Man5 to l, 40 Man8 and of mature N-glycans from the truncated structure a, XM3GN2 to m, bGN3FXM3GN2 (see, Table 3). In addition to the N-glycans characterized in the control callus (Figure 9A), the MALDI-TOF MS of the glycan mixture from selected corn calli expressing GnTIII (Figure 9B) 45 showed at least one ion assigned to N-linked glycans that result from the action of the human GnTIII

enzyme (for a comparison see, Table 3). This oligosaccharide, GN3XM3GN2 (**m**) represents 20 % of the population and contains three GlcNAc residues each linked to one of the three mannoses of the trimannosyl core structure of the N-linked glycan. Analysis of glycan structure through MALDI-TOF as performed here cannot distinguish between GlcNAc residues β (1,2)- or β (1,4)-linked to mannose.

5 Hence it is not clear if or to what extent the structures GN2XM3GN2 (**h**) and GN2FXM3GN2 (**k**) have bisecting oligosaccharides. Both had increased numbers in GnTIII corn cells compared to untransformed control corn cells. Additional experiments are required to reveal that these structures are a mix of normal and bisected oligosaccharides or a single compound.

Besides the new appearance of saccharide structure **m** (bGN3FXM3GN2) in GnTIII corn, it
10 is apparent from the comparison of the glycoforms of control and GnTIII corn, as shown in Table 3, that the amount of structures harbouring high-mannose type N-glycans (M4 and higher) is reduced more than twofold (from 19 % to 7 %) which can be attributed mostly to the reduction of M4-containing N-glycans (from 10 % to 1 % of total) in GnTIII corn versus control corn. In addition the amount of glycoforms having two or more GlcNAc residues has increased from 16 % to 42 %
15 (control versus GnTIII).

In a follow-up experiment, endogenous glycoproteins were isolated from control corn calli and selected corn calli expressing GnTIII based on analysis for the presence of c-myc tag sequence by Western blotting. A comparison of the structures of the N-glycans isolated from glycoproteins present in calli is presented in Table 4. MALDI-TOF allows for the detection of different molecular
20 species in the pool of the N-glycans (glycoforms) and shows a mixture of ions that were assigned to (M+Na)+ adducts of high-mannose (Man)- type N-glycans ranging from **d**, Man5 to 1, Man8 and of mature N-glycans from the truncated structure **a**, XM3GN2 to **k**, GN2FXM3GN2 in control corn.

Remarkably, in transgenic corn expressing GnTIII (Table 4, GnTIII-2), only three isoforms could be detected: FXM3GN2 (**b**; accounting for 9 % of total), GN2FXM3GN2/bGN2FXM3GN2
25 (**k**; 38 %) and bGN3FXM3GN2 (**m**; 54 %). It is not clear if or to what extent the structure depicted as **k** (GN2FXM3GN2/bGN2FXM3GN2) has bisecting oligosaccharides. Its presence is significantly increased in GnTIII corn compared to control corn. Additional experiments are required to reveal that these structures are a mix of normal and bisected oligosaccharides or a single compound.

Besides the new appearance of saccharide structure **m** (bGN3FXM3GN2) in GnTIII corn (54 %), it is apparent from the comparison of the glycoforms of control and GnTIII corn, as summarized in Table 4, that the amount of structures harbouring high-mannose type N-glycans (M4 and higher) is reduced to nil in GnTIII corn versus control corn. Furthermore, the total amount of N-glycans bearing 2 or more (3) GlcNAc residues has increased from 16 to 92% (control versus GnTIII) suggesting that the introduction of bisected GlcNAc residue protects the glycan from degradation by endogenous hexosaminidases as observed before for transgenic GnTIII tobacco.
35

Additioanlly, MALDI-TOF mass spctroscopy data (Figure 11) demonstrate the bisected GlcNAc structure.

Example 14

Oligosaccharide distributions and level of bisected complex oligosaccharides in wildtype and selected transgenic corn plants. Endogenous glycoproteins were isolated from control corn plant leaves and selected corn plant leaves expressing GnTIII based on analysis for the presence of c-myc tag sequence by Western blotting or lectin blotting using E-PHA. A comparison of the structures of the N-glycans isolated from glycoproteins present in leaves is presented in Table 6. MALDI-TOF allows for the detection of different molecular species in the pool of the N-glycans (glycoforms) and shows a mixture of ions that were assigned to (M+Na)⁺ or (M+K)⁺ adducts of high-mannose (Man)-type N-glycans ranging from **f**, Man₆ to **h**, Man₈ and of mature N-glycans from the truncated structure **a**, XM3GN2 to **g**, GN2FXM3GN2 in control corn plant and to **i**, bGN3FXM3GN2 in transgenic GnTIII corn plant.

In addition to the N-glycans characterized in the control plants (Figure A), the MALDI-TOF MS of the glycan mixture from selected corn plant expressing GnTIII (Figure B) showed at least one ion assigned to N-linked glycans that result from the action of the human GnTIII enzyme (for a comparison see Table 6). This oligosaccharide, GN3XM3GN2 (**i**) represents 15% of the population and contains three GlcNAc residues each linked to one of the three mannoses of the trimannosyl core structure of the N-linked glycan. Analysis of glycan structure through MALDI-TOF as performed here cannot distinguish between GlcNAc residues β (1,2)- or β (1,4)-linked to mannose. Hence it is not clear if or to what extent the structure GN2FXM3GN2 (**g**) has bisecting oligosaccharides. It has increased in GnTIII corn compared to control corn plant (23 % versus 5 % in control). Additional experiments are required to reveal that these structures are a mix of normal and bisected oligosaccharides or a single compound. Besides this it is apparent from the comparison of the glycoforms of control and GnTIII corn plants, as depicted in Table 6, that the amount of structures harbouring FXM3GN2 is reduced twofold (from 59 to 30) and the amount of glycoforms having two or more GlcNAc residues has increased from 5 to 38 % (control versus GnTIII).

Additionally, Figure 14 shows a comparison of MALDI-TOF mass spectra of N-glycans of glycoproteins isolated from leaves of control corn (A) and of selected GnTIII-corn plants. GnTIII corn plant was obtained through transformation with human GnTIII gene sequence and selection was performed by Western blotting using either c-myc tag or E-PHA lectin. See Table 6 for structures and abbreviations.

It is understood that the present invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagent, etc., described herein, as these may vary. It is also to be understood that the terminology used herein is used for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention. It must be noted that as used herein and in the appended claims, the singular forms "a," "an" and "the" include plural reference unless the context clearly dictates otherwise.

Unless defined otherwise, all technical and scientific terms used herein have the same

meanings as commonly understood by one of ordinary skill in the art to which this invention belongs.

The invention described and claimed herein is not to be limited in scope by the specific embodiments herein disclosed, since these embodiments are intended as illustrations of several aspects of the invention. Any equivalent embodiments are intended to be within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims.

Various references are cited herein, the disclosures of which are incorporated by reference in their entireties.

Table 1. Structure, molecular weight and percentage of total pool of N-glycans isolated from control and selected GnTIII-17 plants.

Structure	Abbreviation	Name	Mol. Wt.	%	
				Wildtype	GnTIII-17
	XM3GN2	A	1065	8	4
	FM3GN2	B	1079	3	0
	FXM3GN2	C	1211	30	4
	Man5	D	1257	2	3
	GNXM3GN2	E	1268	4	0
	GNFXM3GN2	F	1414	10	2
	Man6	G	1419	3	5
	GN2XM3GN2	H	1471	-	14
	Man7	I	1581	3	4
	GN2FXM3GN2	J	1617	-	
	GN3XM3GN2	K	1674	-	8

	Man8	L	1743	2	4
GlcNAc β 1-2Man α 1 GlcNAc β 1-4Man β 1-4GlcNAc β 1-4GlcNAc GlcNAc β 1-2Man α 1 Xyl β 1 Fuc α 1	GN3FXM3GN2	M	1820	-	31
	Man9	N	1905	1	2

5 **Table 2.** Comparison of the results of mass spec (MALDI-TOF) analysis of N-glycans of endogenous glycoproteins isolated from control tobacco and selected GnTIII-17 plant. See also Table 1.

m/z	Type	Wildtype endo	GnTIII-17 endo
1065	XM3	8	4
1079	FM3	3	0
1211	FXM3	30	4
1257	M5	2	3
1268	GNXM3	4	0
1414	GNFXM3	10	2
1419	M6	3	5
1471	GNbGNXM3		14
1471	GN2XM3	4	
1581	M7	3	4
1617	GN2FXM3	29	19
1617	GNbGNXM3		
1674	GN2bGNXM3	0	8
1743	M8	2	4
1820	GN2bGNFXM3	0	31
1905	M9	1	2
	TOTAL	99	100

Table 3. Overview N-glycans observed in control and transgenic GnTIII corn.

Comparison of N-glycan structures (% of total) found on endogenous glycoproteins of control, untransformed corn and transgenic corn callus expressing GnTIII that could be annotated.

Corresponding mass spectra obtained through MALDI-TOF analyses are given below and
5 saccharides are indicated under column "name." Bisecting GlcNAc residues are depicted as bGN.

Structure abbreviation	m/z	name	Corn callus	
			control	GnTIII
XM3GN2	1065	a	1	5
FXM3GN2	1211	b	37	36
XM4GN2	1227	c	6	1
M5GN2	1257	d	1	1
GNFXM3GN2	1414	e	12	3
M6GN2	1419	f	5	4
GNXM4GN2	1430	g	3	
GN2XM3GN2	1471	h		3
bGN2XM3GN2				
GNFXM4GN2	1576	i	1	
M7GN2	1581	j	1	1
GN2FXM3GN2	1617	k	16	19
bGN2FXM3GN2				
M8GN2	1743	l	1	
bGN3FXM3GN2	1820	m		20
		Total	84	93

Table 4. Schematical overview N-glycans observed in control and transgenic GnTIII corn-2.

Comparison of N-glycan structures (% of total) found on endogenous glycoproteins of control, untransformed corn and transgenic corn callus expressing GnTIII that could be annotated.

Transgenic corn was selected using c-myc tag. Corresponding mass spectra obtained through

5 MALDI-TOF analyses are given below and saccharides are indicated under column "name".

Bisecting GlcNAc residues are depicted as bGN.

Structure abbreviation	m/z	name	Corn callus	
			control	GnTIII-2
XM3GN2	1065	a	1	
FXM3GN2	1211	b	37	9
XM4GN2	1227	c	6	
M5GN2	1257	d	1	
GNFXM3GN2	1414	e	12	
M6GN2	1419	f	5	
GNXM4GN2	1430	g	3	
GN2XM3GN2	1471	h		
bGN2XM3GN2				
GNFXM4GN2	1576	i	1	
M7GN2	1581	j	1	
GN2FXM3GN2	1617	k	16	38
bGN2FXM3GN2				
M8GN2	1743	l	1	
bGN3FXM3GN2	1820	m		54
		Total	84	101

Table 5. Positive test results for E-PHA binding.

Sample ID number	Rating for E-PHA Binding	Included in Pooled positive sample
1.	Unclear	
2.	Unclear	
3.	1	
4.	0	
5.	0	
6.	0	
7.	0	
8.	0	
9.	0	
10.	Unclear	
11.	Unclear	
12.	1	
13.	1	
14.	0	
15.	0	
16.	0	
17.	0	
18.	0	
19.	0	
20.	0	
21.	0	
22.	0	
23.	1	
24.	1	
25.	3	Yes
26.	3	Yes
27.	0	
28.	0	
29.	0	
30.	0	
31.	0	
32.	0	
33.	2	Yes
34.	1	
35.	0	
36.	0	
37.	Unclear	
38.	1	
39.	0	
40.	1	
41.	1	
42.	0	
43.	0	
44.	Unclear	
45.	0	
46.	0	

-59-

47.	0	
48.	2	Yes
49.	0	
50.	0	
51.	0	
52.	0	
53.	0	
54.	1	
55.	2	Yes
56.	2	Yes
57.	2	Yes
58.	1	
59.	2	Yes
60.	0	
61.	0	
62.	0	
63.	0	
64.	0	
65.	0	
66.	0	
67.	0	
68.	0	
69.	Negative control	

Table 6. Schematical overview N-glycans observed in control and transgenic GnTIII corn plants.

Structure abbreviation	m/z	name	Corn plant	
			control	GnTIII
XM3GN2	1065	a	4	14
FM3GN2	1079	b	2	
FXM3GN2	1211	c	59	30
XM4GN2	1227	d	3	12
GNFXM3GN2	1414	e	10	6
M6GN2	1419	f	2	
GN2FXM3GN2	1617	g	5	23
bGN2FXM3GN2				
M8GN2	1743	h	1	
bGN3FXM3GN2	1820	i		15
			Total	86
				100

CLAIMS

What is Claimed is:

- 5 1. A plant host cell system, comprising a mammalian UDP-N-acetylglucosamine: β -D mannoside β (1,4)-N-acetylglucosaminyltransferase (GnTIII) enzyme.
2. The plant host system according to Claim 1, wherein said GnTIII is a human GnTIII.
- 10 3. The plant host system according to Claim 1, wherein said system is a portion of a plant.
4. The plant host system according to Claim 1, wherein said system is a portion of a plant selected from the group consisting of a cell, leaf, embryo, callus, stem, pericarp, protoplast, root, tuber, kernel, endosperm and embryo.
- 15 5. The plant host system according to Claim 1, wherein said system is a whole plant.
6. The plant host system according to Claim 1, further comprising a heterologous glycoprotein.
- 20 7. The plant host system according to Claim 5, wherein said heterologous glycoprotein protein comprises an antibody or fragment thereof.
8. The plant host system according to Claim 5, wherein said heterologous glycoprotein or functional fragment thereof comprises bisected oligosaccharides.
- 25 9. The plant host system according to Claim 5, wherein said heterologous glycoprotein comprises bisected glycans with galactose residues.
10. The plant host system according to Claim 1, wherein said plant is a tobacco plant.
- 30 11. The plant host system according to Claim 1, which further comprises a functional protein selected from the group consisting of a transporter and an enzyme providing N-glycan biosynthesis.
- 35 12. The plant host system according to Claim 11, wherein said enzyme is a (human) β -1,4 galactosyltransferase.
13. The plant host system according to Claim 11, which further comprises a heterologous glycoprotein, having an increased number of galactose residues.

14. A plant host system comprising a nucleic acid sequence encoding a mammalian GnTIII protein.
15. A plant host system comprising a vector comprising a nucleic acid sequence encoding a mammalian GnTIII protein.
5
16. The plant host system according to Claim 15, which further comprises a nucleic acid sequence encoding a functional protein selected from a group consisting of a transporter and an enzyme providing N-glycan biosynthesis.
10
17. A method comprising a) crossing a plant expressing a heterologous glycoprotein with a plant according to Claim 5, b) harvesting progeny from said crossing and c) selecting a desired progeny plant.
15
18. The method according to Claim 17, wherein said desired progeny plant expressing said heterologous glycoprotein protein having bisected oligosaccharides.
19. The method according to Claim 17, wherein said plant host system is a transgenic plant.
20. A method for obtaining a heterologous glycoprotein having bisected oligosaccharides comprising introducing a nucleic acid sequence encoding GnTIII that is normally not present in plant into a plant host system and a nucleic acid sequence encoding a heterologous glycoprotein and isolating said heterologous glycoprotein.
25
21. The method according to Claim 20, wherein said nucleic acid sequences are introduced into a plant cell and said plant cell is regenerated into a plant.
22. The method according to Claim 20, wherein said nucleic acid sequences are introduced into a plant host system by transforming said plant host system with a vector comprising a acid sequence encoding GnTIII that is normally not present in plant into a plant and a nucleic acid sequence encoding a heterologous glycoprotein.
30
23. The method according to Claim 20, wherein said nucleic acid sequences are introduced into a plant host system by transforming said plant host system with a vector comprising a nucleic acid sequence encoding GnTIII that is normally not present in plant into a plant and a nucleic acid sequence encoding a heterologous glycoprotein.
35
24. The method according to Claim 20, wherein said nucleic acid sequences are introduced into a plant host system by transforming said plant with a vector comprising a nucleic acid sequence

encoding GnTIII that is normally not present in plant into a plant host system and vector comprising a nucleic acid sequence encoding a heterologous glycoprotein.

25. A method for obtaining a heterologous glycoprotein having bisected oligosaccharides
5 comprising cultivating the regenerated plant obtained in claim 21.
26. A method for obtaining a desired glycoprotein comprising a) cultivating the plant host system of
Claim 6 and b) harvesting and fractionating said plant.
- 10 27. A plant obtainable by a method according to Claim 19.
28. A method for obtaining a plant host system comprising a) crossing a plant comprising a
functional protein selected from a group consisting of a transporter or an enzyme providing N-
glycan biosynthesis with a plant according to Claim 5, b) harvesting progeny from said crossing
15 and c) selecting a desired progeny.
29. A transgenic plant obtained according to the method of Claim 28.
30. A method for increasing galactosylation of a heterologous glycoprotein expressed in a plant host
20 system comprising a) introducing a nucleic acid sequence encoding GnTIII and a sequence
selected from a group consisting of a transporter or an enzyme not normally present in a plant
into said plant host system expressing said heterologous glycoprotein and b) isolating said
glycoprotein.
- 25 31. A plant derived glycoprotein comprising bisected oligosaccharides.
32. Use of a plant host system according to Claims 1-16 to produce a desired glycoprotein or
functional fragment thereof.
- 30 33. Use according to Claim 32 wherein a said glycoprotein or functional fragment thereof comprises
bisected oligosaccharides.
34. A plant-derived glycoprotein or functional fragment thereof obtained by a method according to
Claim 19.
- 35 35. Use of a glycoprotein or functional fragment thereof according to Claim 31 for the production of
a pharmaceutical composition.
36. A composition comprising a glycoprotein or functional fragment thereof according to Claim 31.

37. An isolated hybrid protein comprising an active site of GnTIII and a transmembrane region of a protein, said protein residing in endoplasmic reticulum or Golgi apparatus of a eukaryotic cell.

5 38. The protein according to Claim 37, wherein said protein residing in endoplasmic reticulum or Golgi apparatus of a eukaryotic cell is an enzyme.

39. The protein according to Claim 37, wherein said protein residing in endoplasmic reticulum or Golgi apparatus of a eukaryotic cell is a glycosyltransferase.

10 40. The protein according to Claim 37, wherein said protein residing in endoplasmic reticulum or Golgi apparatus of a eukaryotic cell is a glycosyltransferase selected from the group consisting of a mannosidaseI, mannosidaseII, GnTI, GnTII, Xy1T and FucT,

15 41. The protein according to Claim 37, wherein said protein residing in endoplasmic reticulum or Golgi apparatus of a eukaryotic cell is a plant protein.

42. An isolated nucleic acid sequence encoding the protein of Claim 31.

20 43. A vector comprising the isolated nucleic acid sequence of Claim 42.

44. A plant comprising the isolated nucleic acid sequence of Claim 42.

25 45. The plant according to Claim 44 which further comprises a nucleic acid sequence encoding a heterologous glycoprotein.

30 46. A method for providing a transgenic plant capable of expressing a heterologous glycoprotein with the capacity to extend an N-linked glycan with galactose comprising crossing a transgenic plant with a plant according to Claim 44, harvesting progeny from said crossing and selecting a desired progeny plant expressing said recombinant protein and expressing a functional (mammalian) enzyme involved in (mammalian) N-glycan biosynthesis that is normally not present in plants.

35 47. A method for providing a transgenic plant capable of expressing a heterologous glycoprotein with the capacity to extend an N-linked glycan with galactose comprising introducing the nucleic acid sequence of Claim 42 and a nucleic acid sequence encoding said heterologous glycoprotein.

48. A method, comprising:

- a. providing: i) a plant cell, and ii) an expression vector comprising nucleic acid encoding a GNTIII enzyme; and
- b. introducing said expression vector into said plant cell under conditions such that said enzyme is expressed.

5

49. The method of Claim 22, wherein said nucleic acid encoding a GNTIII comprises the nucleic acid sequence of SEQ ID NO:1.

50. A method, comprising:

- 10 a. providing: i) a plant cell, ii) a first expression vector comprising nucleic acid encoding a GNTIII enzyme, and iii) a second expression vector comprising nucleic acid encoding a heterologous glycoprotein; and
- b. introducing said first and second expression vectors into said plant cell under conditions such that said hybrid enzyme and said heterologous protein are expressed.

15

51. The method of Claim 50, wherein said heterologous protein is an antibody or antibody fragment.

52. A method, comprising:

- 20 a) providing: i) a first plant comprising a first expression vector, said first vector comprising nucleic acid encoding a GNTIII enzyme, and ii) a second plant comprising a second expression vector, said second vector comprising nucleic acid encoding a heterologous protein; and
- b) crossing said first plant and said second plant to produce progeny expressing said hybrid enzyme and said heterologous protein.

25

53. A plant, comprising first and second expression vectors, said first vector comprising nucleic acid encoding a GNTIII enzyme, said second vector comprising nucleic acid encoding a heterologous protein.

30 54. The plant of Claim 53, wherein said heterologous protein is an antibody or antibody fragment.

FIG. 1A

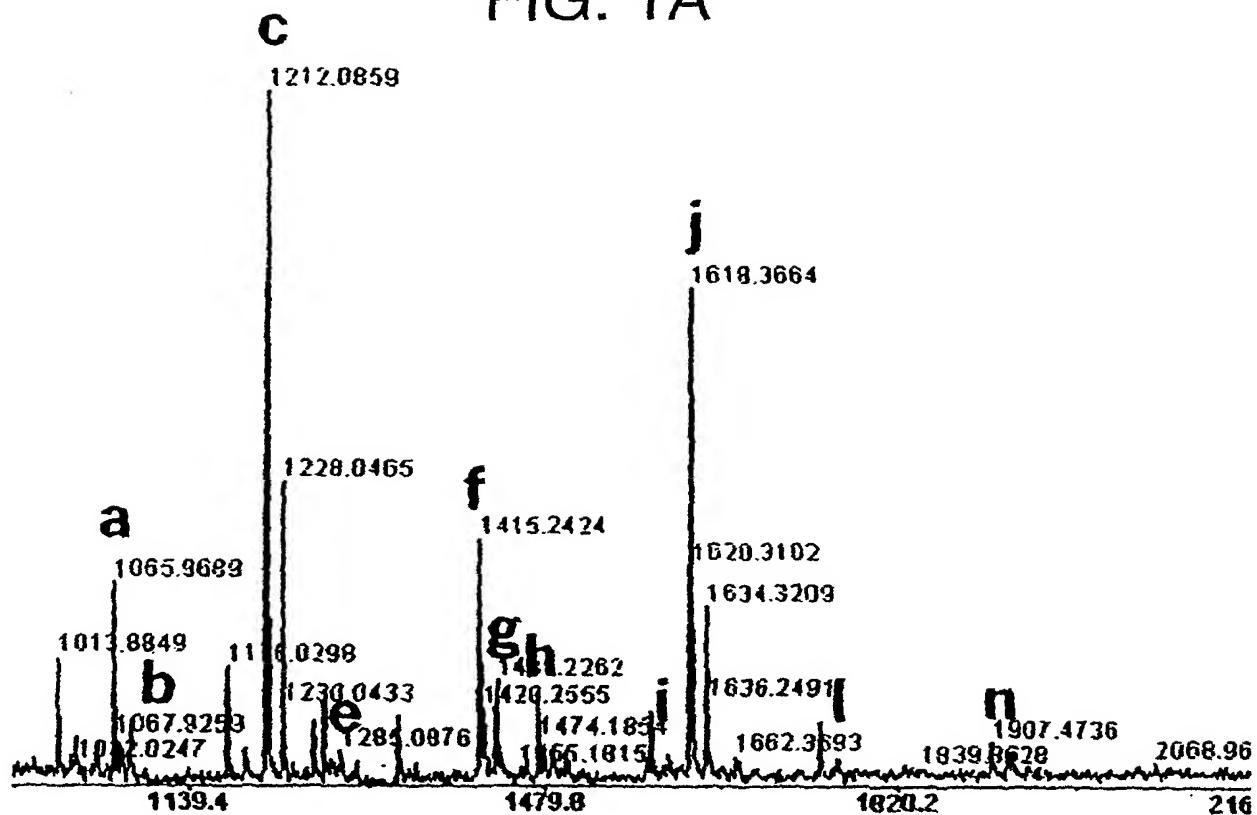
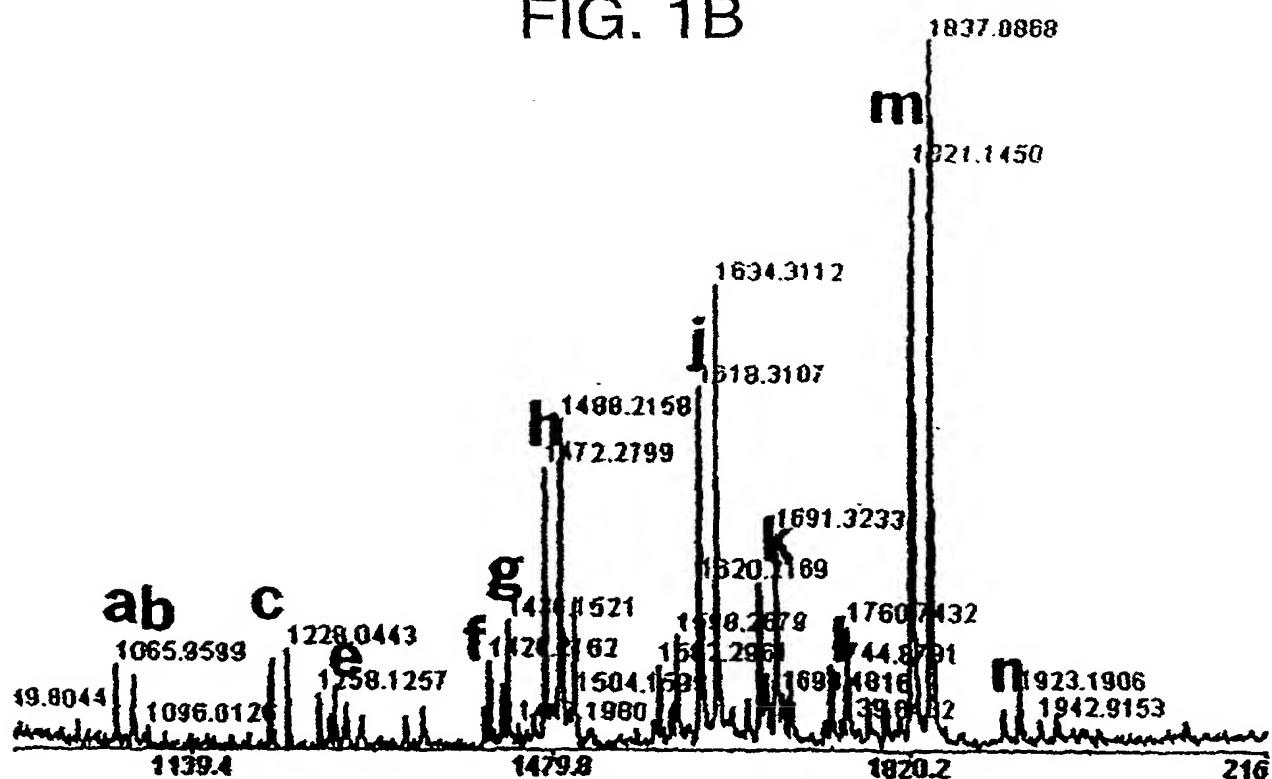


FIG. 1B



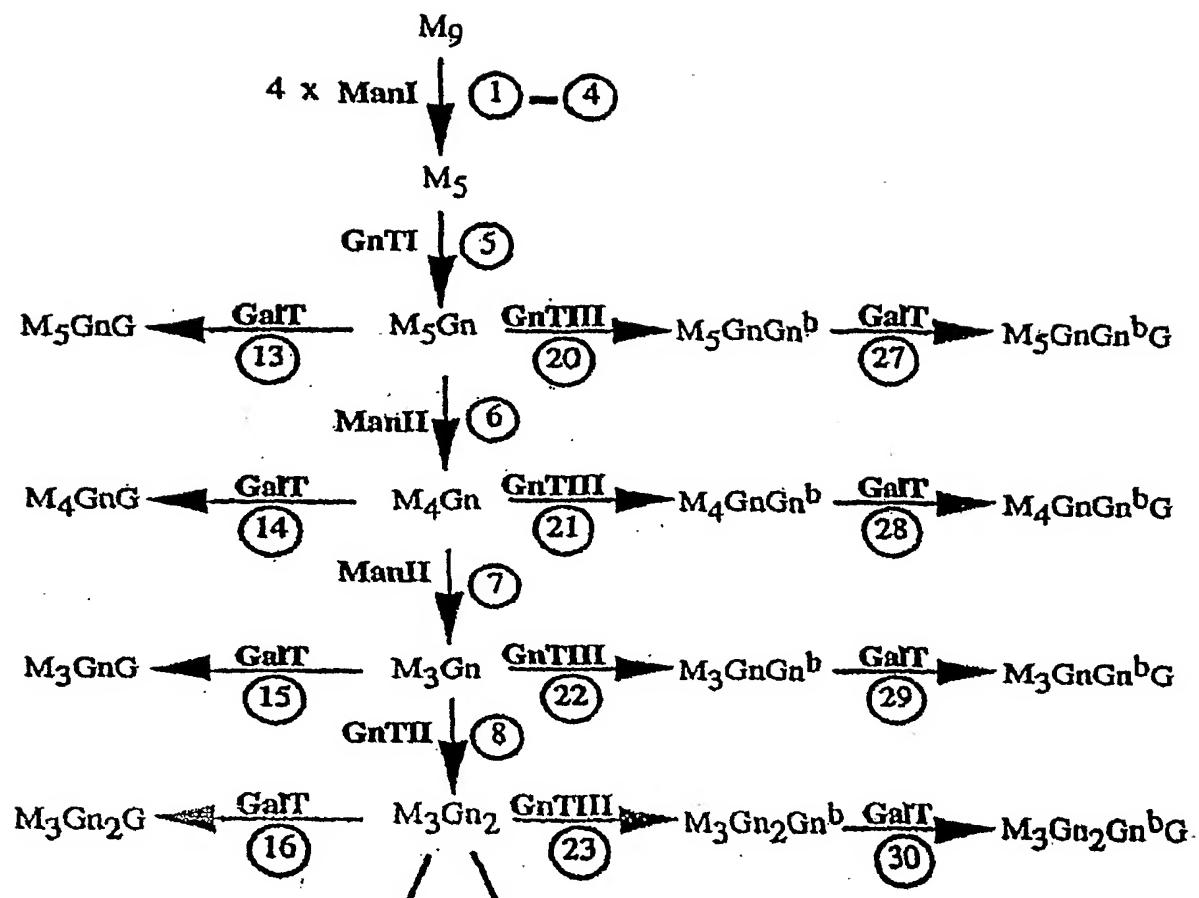


FIG. 2

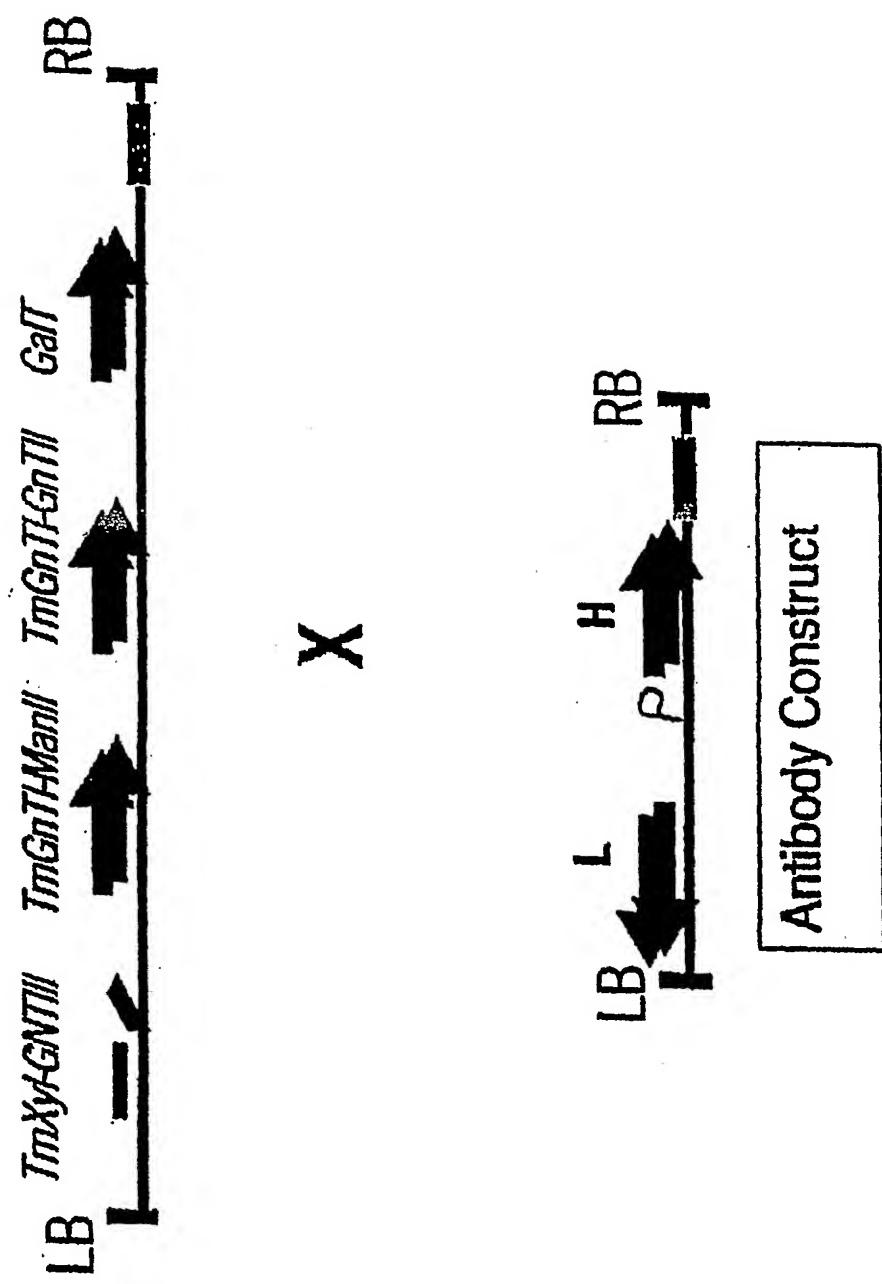


FIG. 3

Nucleotide sequence (incl. myc-tag)

CCATGGTGATGAGACGCTACAAGCTTTCTCATGTTCTGTATGGCCGGCCTG
TGCCTCATCTCCTCCTGCACTTCTCAAGACCCGTGCCTATGTCACCTTCCCC
CGAGAACTGGCCTCCCTCAGCCCTAACCTGGTGTCCAGCTTTCTGGAACAA
TGCCCGGTACGCCAGGCCAGCCCCGAGCCAGGAGGCCGTGACCTGCTG
CGTACCCCACACTACTCCCACCGCCCTGCTGCAGCCGCTGCCGCCAGCAA
GGCGGCCGAGGAGCTCACCGGGTGGACTGGTGTGCCAGGACACCACC
GAGTATTTCGTGCGCACCAAGGCCGGCGTCTGCTTCAAACCCGGCACCA
AGATGCTGGAGAGGCCGCCGGGACGCCGGAGGAGAACCTGAGGGGG
CCAACGGCTCCTCGGCCGGCGGCCACCCGGTACCTCCTGAGCGCCCGGGA
GCGCACGGGGGGCGAGGCCGGCGCAAGTGGGTGGAGTGCCTGCGCT
GCCCGGCTGGCACGGACCCAGCTGCCGTGCCACTGTGGTGCAGTACTCC
AACCTGCCACCAAGGAGCGGCTGGTGCCAGGGAGGTGCCGCCGTCA
TCAACGCCATCAACGTCAACCACGAGTTGACCTGCTGGACGTGCGCTTCCA
CGAGCTGGCGACGTGGTGGACGCCCTTGTTGAGTCCAACTTCACG
GCTTATGGGAGCCGCCGCTCAAGTCCGGAGATGCTGACCAATGGCA
CCTCGAGTACATCCGCCACAAGGTGCTATGTCCTGGACCACCTCCG
CCCGGCCGGCCGGCAGGACGGCTGGATGCCGACGACTACCTGCGCACCTCC
TCACCCAGGACGGCGTCTCGCGCTGCCAACCTGCCGCCGACGACGTCTT
CATCATTGACGATGCCGACGAGATCCCGCCCGTGACGGCGTCCTTCTCA
AGCTCTACGATGGCTGGACCGAGCCCTCGCCTTCCACATGCGCAAGTCGCTC
TACGGCTTCTTCTGGAAGCAGCCGGCACCCCTGGAGGTGGTGTCAAGGCTGCA
CGGTGGACATGCTGCAGGCAGTGTATGGCTGGACGGCATCCGCTGCCCG
CCGCCAGTACTACACCATGCCAACCTCAGACAGTATGAGAACCGCACCGGC
CACATCCTGGTGCAGTGGTCGCTGGCAGCCCCCTGCACCTGCCGGCTGGC
ACTGCTCCTGGTGTTCACGCCAGGGCATCTACTCAAGCTCGTGTCCGCC
CAGAATGGCGACTTCCCACGCTGGGTGACTACGAGGACAAGCAGGGACCTGA
ACTACATCCGCCCTGATCCGACCCGGGCTGGTTCGACGGCACCGCAGCA
GGAGTACCCGCCCTGCAGACCCAGCGAGCACATGTATGCCCAAGTACCTG
CTGAAGAACTACGACCGGTTCCACTACCTGCTGGACAACCCCTACCAGGAGC
CCAGGAGCACGGCGGGCGGGTGGGCCACAGGGTCCCAGGGAAAGGC
CGCCCGCCCGGGCAAACGGACGAGGCCAGTCAACAAAAACTCATCT
CAGAAGAGGATCTGAATTAGGATCC

FIG. 4A

PROTEIN SEQUENCE

MVMRRYKFL MFCMAGLCLI SFLHFFKTL S YVTFPRELAS LSPNLVSSFF
WNNAPVTPQA SPEPGGPDLL RTPLYSHSPL LOPLPPSKAA EELHRVDLVL
PEDTTEYFVR TKAGGVCFKP GTKMLERPPP GRPEEKPEGA NGSSARRPPR
YLLSARERTG GRGARRKWVE CVCLPGWHGP SCGVPTVVQY SNLPTKERLV
PREVPRRVIN AINVNHEFDL LDVRFHELGD VVDAFVVCES NFTAYGEPRP
LKFREMLTNG TFYEYIRHKVL YVFLDHFPNG GRODGWIADD YLRTFLTODG
VSRLRNLRPD DVFIIDDADE IPARDGVLF KLYDGWTEPF AFHMRKSILYG
FFWKQPGTLE VVSGCTVDMQ QAVYGLDGIR LRRQYYTMP NFROYENRTG
HILVQWSLGS PLHFAGWHCS WCFTPEGIYF KLVSQONGDF PRWGDYEDKR
DLNYIRGLIR TGGWFDGTOQ EYPPADPSEH MYAPKYLLKN YDRFHILLDN
PYQEPRSTAA GGWRHRGPEG RPPARGKLDE AEVEQKLISE EDLN

FIG. 4B

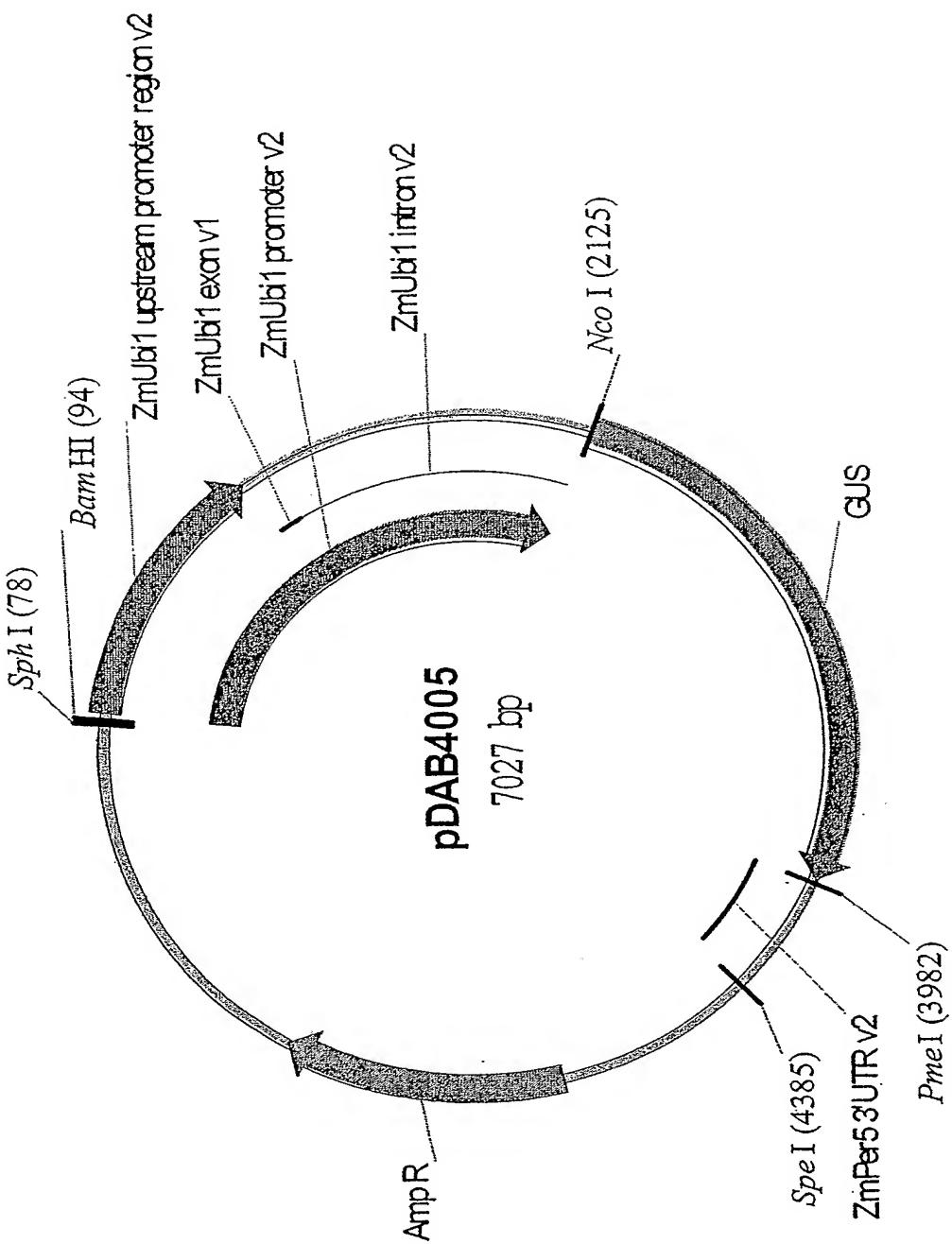


FIG. 5A

Plasmid pDAB4005 (7027 bp)
 Description: ZmUb1 promoter + intron/GUS/Per5 3'UTR

Nucleotide start	Nucleotide end	Sequence Feature
1	115	Linker sequence with multiple cloning site: CATGATTACGCCAAGCTAGCGGCCATTCCGGAAAGCTA GGCCACCCTGGCCCGCCTGCAGGGAAAGCTTGCATGCCCTGC AGATCCCCGGGATCCTCTAGAGTCGACCTGCA
116	2105	Maize ubiquitin promoter + intron
1093	2105	Maize ubiquitin intron
2106	2125	Linker sequence: GGGTACCCCCGGGGTGCAC
2126	3967	GUS coding region
3968	3985	Linker sequence: TAATGAGCTCGTTAAA
3986	4350	Maize peroxidase-5 3' UTR
4351	4405	Linker sequence: CGGGCGCCCTAGCTAGCCZACGGTGCCCCAGATCCACTAGTTC TAGAGCGGCCGCTT
4406	7027	Puc19
5006	5866	Ampicillin resistance gene

Sequence

1 CATGATTACG CCAAGCTAGC GGCGCATT CCGGAAGCT AGGCCACCGT
 51 GGCCCGCTG CAGGGAAAGC TTGCATGCC GCAGATCCCC GGGGATCCTC
 101 TAGAGTCGAC CTGGCACTGCA GCGTGACCCG GTCTGTCCCC TCTCTAGAGA

FIG 5B

151 TAATGAGCAT TGCATGTCTA AGTTATAAAA ATTACCCACA TATTTTTTT
 201 GTCACACTTG TTGAAAGTGC AGTTATCTA TCTTTATACA TATTTTTAAA
 251 CTTTAATCTA CGAATAATAT AATCTATAGT ACTACAATAA TATCAGTGT
 301 TTAGAGAACAT ATATAAATGA ACAGTAGAC ATGGTCTZAA GGACAATTGA
 351 GTATTTGAC AACAGGACTC TACAGTTTA TCTTTTGT GTGCATGTGT
 401 TCTCCTTTT TTGCAAAAT AGCTTACCT ATATAATACT TCATCCATT
 451 TATTAGTACA TCCATTAGG GTTTAGGGTT ATCTATTT AGCCTCTAAA TTAGAAAAC
 501 TTTTTAGTA CATCTATTT ATTCTATTT AGCCTCTAAA TTAGAAAAC
 551 TAAAACCTA TTATTAGTTT TTATTAAAT ATTCTAGATA TAAAATAGAA
 601 TAAAATAAAG TGACTAAAAA TTAAACAAAT ACCCTTTAAG AAATTAAAAA
 651 AACTAAGGAA ACATTTTCT TGTTTCGAGT AGATAATGCC AGCCTGTAA
 701 ACGCCGTGCA CGAAGCTAAC GGACACCAAC CAGCGAACCA GCAGCGTCGC
 751 GTCGGGCAA GCGAAGCAGA CGGCACGGCA TCTCTGTGCG TGCTCTGGA
 801 CCCCTCTCGA GAGTTCGGCT CCACCGTTGG ACTTGCTCCG CTGTCGGCAT
 851 CCAGAAATIG CGTGGGGAG CGGCAGACGT GAGCCGGCAC GGAAGGGGGC
 901 CTCCTCTCC TCTCACGGCA CGGCAGCTAC GGGGATTC TTTCCCACCG
 951 CTCCTCGCT TTCCCTCCCT CGCCCGCCGT AATAAATAGA CACCCCTCC
 1001 ACACCCCTT TCCCCAACTT CGTGTGTTC GGAGCGACACA
 1051 CAGATCTCC CCAAATCCAC CGGTGGCAC CTCGCTTCAGTC AGGTACGGCC
 1101 CTCGTCCTCC CCCCCCCCCC CTCTCTACCT TCTCTAGATC GGCGTTCCGG
 1151 TCCATGCATG GTAGGGCCC GGTAGTTCTA CTTCGTGTCAG TGTGTGTGTT
 1201 AGATCCGTGT TTGTGTAGA TCCGTTGCTGC TAGCCTGTGT ACACGGATGC
 1251 GACCTGTACG TCAAGACGT TCTGATTGCT AACTTGCCAG TGTTCCTCTT
 1301 TGGGAAATCC TGGGATGGCT CTAGCCGTTTC CGCAGACGGG ATCGATTCA
 1351 TGATTTTTT TGTTTCGTTG CATAGGGTTT GGTTTGCCT TTTCTTTAT
 1401 TCAATATAAT GCGGTGCACT TGTTTGTGCG GTCATCTTTT CATGCTTTT
 1451 TTGTCCTGG TTGTGTGATGAT GTGGTCTGGT TGGGGTTCAG TTCTAGATCG
 1501 GAGTAGAATT CTGTTCAA CTACCTGGTG GATTATIAA TTITGGATCT
 1551 GTATGTTGTGT GCCATACATA TTCAATAGTTA CGAATTGAAG ATGATGGATG
 1601 GAAATATCGA TCTAGGATAG GTATAACATGT TGATGCGGGT TTIACATGATG
 1651 CATAACAGA GATGCTTTT GTTCGCTTGG TTGTGATGAT GTGGTGTGGT

FIG 5B CONT.

1701 TGGGGGGTCG TTCATTCGT CTAGATCGGA GTAGAACTACT GTTCAAACT
 1751 ACCTGGTGTAA TTTATAATT TTGGAACCTGT ATGGTGTGTGT CATACTATCTT
 1801 CATACTAACG AGTTTAAGAT GGATGGAAT ATCGCATCTAG GATAGGTATA
 1851 CATTGGTGTAG TGGGTTTAC TGATGCATAT ACATGATGGC ATATGAGCA
 1901 TCTATTCTA TGCTCTAACCG TTGAGTACCT ATCTATTATA ATAACAAGT
 1951 ATGGTTTATA ATTATTTGA TCTTGTATATA CTGGATGAT GGCATATGCC
 2001 GCAGCTTAT GTGGATTTTT TTAGCCCTGCG CTTCATACGC TATTATTTG
 2051 CTTGGTACTG TTTCTTTTG CGATGTCAC CTTGGTGTGTT GGTGTTACTT
 2101 CTGCAGGGTA CCCCGGGGT CGACCATGGT AAGGGGAGC CACCACCC
 2151 ACCACCACAT GGTCGGTCT GTAGAAACCC CAACCCGTGA AATCAAAAAA
 2201 CTCGACGGCC TGTGGGCAT CAGTCTGGAT CGCGAAAATC GTGGAATTGA
 2251 TCAGCGTGG TGGAAGGC CGTTACAAGA AAGCGGGCA ATTGCTGTGC
 2301 CAGGCAGTT TAACGATCAG TTGCGCGATG CAGATATTG TAATTATGCC
 2351 GGCAACGGTCT GGTATCAGCG CGAAGTCTTT ATACCGAAG GTTGGCGAGG
 2401 CCAGCGTATC GTGCTGGCTT TCGATGGGT CACTCATAC GGCAAAAGTGT
 2451 GGGTCAAATAA TCAGGAAGTG ATGGAGCATT AGGGGGCTA TACGCCATT
 2501 GAAGCCCGATG TCACGCCGTAA TGTTATTGCC GGGAAAAGTG TACGTATCAC
 2551 CGTTTGTGTG AACAAACGAAC TGAACCTGGCA GACTATCCCCG CGGGGAATGG
 2601 TGATTACCGA CGAAAACGGC AAGAAAAGC AGTCTTACTT CCATGATTTC
 2651 TTAACTATG CCGGAATCCA TCGCAGCGTA ATGCTCTACA CCACGCCGAA
 2701 CACCTGGGTG GACGGATATCA CGGTGTCAC GCATGTCGGC CAAAGACTGTA
 2751 ACCACGGTC TTGTTGACTGG CAGGGGTGG CCAATGGTGA TGTCAAGCGTT
 2801 GAACTGGGTG ATGGGGATCA ACAGGTGGTT GCAACTGGAC AAGGCACTAG
 2851 CGGGACTTTG CAAGTGGTGA ATCCGCACCT CTGGCAACCG GGTGAAGGTT
 2901 ATCTCTATGA ACTGTGGCTC ACAGGCCAAA GCCAGACAGA GTGGTATATC
 2951 TACCCGCTTC GCCTGGGCAAT CCGGTCACTG GCAACTGGAC AGGGACAGTT
 3001 CCTGATTAAC CACAAACCGT TCTACTTAC TGCTTGGT CGTCATGAAG
 3051 ATGGGACTT ACGTGGCAA GGATTCGATA ACCTGCTGAT GGTGCACGAC
 3101 CACGCATTAA TGACTGGAT TGGGGCAAC TCCTACCGTA CCTCGCATTA
 3151 CCCTTAAGCT GAAAGAGATGC TCGACTGGC AGATGAACAT GGCATCGTGG
 3201 TGATTGATGA AACTGCTGCT GTCGGCTTTA ACCTCTCTTT AGGCATTGGT

FIG 5B CONT.

3251 TTCGAAGGG GCAACAAGCC GAAAGGAACTG TACAGCGAAG AGGCAGTCAA
 3301 CGGGAAACT CAGCAAGGCC ACTTACAGGC GTGGTGTATG GAGTAAAGAG CTGATAGCGC
 3351 GTGACAAAAA CCACCCAAGC CGCAAGTGC ACGGAAATT TGCCCACTGG CAACGAACCG
 3401 GATACCCCTC CGGAAAGTGC CGGGAAATT TGCCCACTGG CGGAAGAAC
 3451 GCGTAAGACT GACCCGACGC GTCCGATCAC CTGGTCAAT GTATGTTCT
 3501 GCGACGCTCA CACCGATAAC ATCAGCGATC TCTTGATGT GCTGTGCCTG
 3551 AACCGTTATT ACGGATGGTA TGTCCAAGC GCGGATTGG AAACGGCAGA
 3601 GAAGGTACTG GAAAAGAAC TTCTGGCCTG GCAGGAGAAA CTGCATCAGC
 3651 CGATTATCAT CACCGAATAC GGGGTGGATA CGTAGCCCC GCTGCACTCA
 3701 ATGTACACCG ACATGTGGAG TGAAGGATAT CAGTGTGCAT GGCTGGATAT
 3751 GTATCACCGC GTCTTGTATC GGGTCAAGGCC CGTCGTGGT GAACAGGTAT
 3801 GGAATTTCGC CGATTTTGCG ACCTCGCAAG GCAATATGG CGTITGGGGT
 3851 ACAAGZAAG GGATCTTAC TCGGGAACCGC AAACCGAAGT CGGGGGCTT
 3901 TCTGCTGCCA AAACGCTGGA CTGGCATGAA CTTCGGTGAA AAACCGCAGC
 3951 AGGGAGGCAA ACAATGATAA TGAGCTCGTT TAAACTGAGG GCACGTGAAGT
 4001 CGCTTGATGT GCTGAATTGT TTGTGATGTT GGTGGGTAT TTGTTTAA
 4051 TAAGTAAGCA TGGCTGTGAT TTATCATAT GATCGATCTT TGGGGTTTA
 4101 TTTAACACAT TGTAATATGT GTATCTTATA ATAACCTAAAT GTATAAGATG
 4151 TGTTCAATTCT TCGGTTGCCA TAGATCTGCT TATTGACCT GTGATGTTT
 4201 GACTCCAAA ACCAAAATCA CAACTCAATA FACTCATGGG ATAATGTCAC
 4251 CTGTTTCCTTG AAGAGTTCAT CTACCAATTCC AGTTGGCAATT TATCAGTGT
 4301 GCAGCGGCC TGTGCTTTGT AACATACAA TTGTTACGGC ATATATCCAA
 4351 CGGGCGGCCT AGCTAGCCAC GTGGGCCAGA TCCACTAGT CTAGAGCGGC
 4401 CGCTTAATTCT ACTGGCCGTC GTTTTACAAC GTCGTGACTG GGAAACCCCT
 4451 GCGGTTACCC AACTTAATCG CCTTGAGCA CATCCCCCT TCGCCAGCTG
 4501 GCGTAATAGC GAAGAGGCC GCACCGATCG CCCTTCCCAA CAGTTGGCA
 4551 GCCTGAAATGG CGAATGGGC CTGATGGGGT ATTTCCTCTT TACGCATCTG
 4601 TGCGGTATT CACACCGCAT ATGGTGCCT CTCAGTACAA TCIGCTCTGA
 4651 TGCGGCAATAG TTAAGCCAGC CCCGACACCC GCCAACCCCC GCTGACGGGC
 4701 CCTGACGGGC TTGTCTGCTC CCGGCATCCG CTTACAGACA AGCTGTGACC
 4751 GTCTCCGGGA GCTGCAATGTG TCAGAGGTTT TCACCGTCAAT CACCGAAACG

FIG 5B CONT.

4801 CGCGAGACGA AAGGGCCTCG TGATACGCC ATTTCATAG GTTAATGTCA
 4851 TGATAATAAT GGTTCTTAG ACGTCAAGGTG GCACITTCG GGGAAATGTG
 4901 CGCGGAACCC CTATTGTTT ATTTCCTAA ATACATTAA ATATGTATCC
 4951 GCTCATGAGA CAATAACCTT GATAAATGCT TCAATAATAT TGAAAAGGA
 5001 AGAGTATGAG TATTCAACAT TTCCGTGTCG CCCTTATTC CTTTTTGCG
 5051 GCATTGGCC TTCCTGTTT TGCTCACCCA GAAACGCTGG TGAAGTAAA
 5101 AGATGGTGA GATCAGTGG GTGCACGAGT GGTTACATC GAACGGATC
 5151 TCAACAGCGG TAAGATCCTT GAGAGTTTC GCCCGAAGA ACGTTTCCA
 5201 ATGATGAGCA CTTTAAAGT TCTGCTATGT GCGCGGGTAT TATCCCGTAT
 5251 TGACGCCGG CAAGGCAAC TCGGTGCGCG CATAACATAT TCTCAGAATG
 5301 ACTTGGTGA GTACTCACCA GTCAACAGAAA AGCATCTTAC GGATGGCATG
 5351 ACAGTAAGAG AATTATGCCAG TGCTGCCATA ACCATGAGTG ATTACACTGC
 5401 GGCCAACCTTA CTTCTGACAA CGATCGAGG ACCGAAGGAG CTAACCGCTT
 5451 TTTTGACAA CATGGGGAT CATGTAACCT GCCTTGATCG TTGGGAACCG
 5501 GAGGTGAATG AAGCCATACCC AAACGACGG CGTGACACCA CGATGCCGT
 5551 AGCAATGGCA ACAACGTTGC GCAAACCTATT ACTGGGCAA CTACTTACTC
 5601 TAGCTTCCCG GCAACAATTAA ATAGACTGGA TGGAGGGGA TAAAGTTGCA
 5651 GGACCACTTC TGCGGCTGGC CCTTCGGCT GGCTGGTTA TTGCTGATAA
 5701 ATCTGGGCC GGTGAGCGTG GGCTCTCCGGT ATCGTAGTTA TCTACACGGC
 5751 CAGATGGTAA GGCCTCCGGT TAATTAAAA GATCTAGGT GTGCCTCACT
 5801 GCAACTATGG ATGAAACGAAA TAGACAGATC GCTGAGATAG TATCATTGCA
 5851 GATTAAGCAT TGGTAACGT CAGACCAAGT TTACTCATAT ATACCTTAGA
 5901 TTGATTAAA ACTTCATTAT TAATTAAAA GATCTAGGT GAAGATCCTT
 5951 TTGATAATC TCATGACCAA ATCCCTTAA CGTGAGTTT CGTTCCACTG
 6001 AGCGTCAGAC CCCGTGAAA AGATCAAAGG ATCTTCTTGA GATCCCTTT
 6051 TTCTGGCGGT AATCTGCTGC TTGCAACAA AAAAACACC GCTACCGCG
 6101 GTGGTTTGGT TGCGGGATA AGAGCTACCA ACTCTTTTC CGAAGGTAAC
 6151 TGGCTTCAGC AGAGGGCAGA TACCAAATAC TGTCTCTCTA GTGTAGCCGT
 6201 AGTTAGGCCA CCACTTCAAG AACTCTGTAG CACCGCCTAC ATACCTCGCT
 6251 CTGCTAATCC TGTTTACCGT GGCTGCTGCC AGTGGCGATA AGTGTGTCT
 6301 TACCGGGTTG GACTCAAGAC GATAAGTACC GGATAAGGCG CAGCGGTGCG

FIG 5B CONT.

6351 GCTGAACGGG GGGTCGTGC ACACAGCCC GCTTGGACCG AACGACCTAC
6401 ACCGAACTGA GATACTACA GCGTGAGCAT TGAGAAAGCG CCAAGCTTCC
6451 CGAAGGGAGA AAGGGGGACA GGTATCCGGT AAGGGGCAAG GTCGGAACAG
6501 GAGAGGCCAC GAGGGAGCTT CCAGGGGAA AGCCCTGGTA TCTTTATAGT
6551 CCTGTGGGT TTGCCCCACTT CTGACTTGAG CGTCGATT TGATGCTC
6601 GTCAGGGGG CGGAGCCAT GGAAAAACGC CAGCAACGGG GCCTTTTAC
6651 GGTTCCTGGC CTTTGGCTGG CCTTTGGCTC ACATGTTCTT TCCTGGTTA
6701 TCCCCTGATT CTGTGGATAA CGGTATTACC GCCTTTGAGT GAGGTGATAAC
6751 CGCTCGCCGC AGGGAAAGA CCGAGCCAG CGAGTCAGTG AGCGAGGAAG
6801 CGGAAGAGCG CCCAATACGC AAACCGCCCT ICCTCGCCG TTGGCCGATT
6851 CATTAAATGCA GCTGGCACGA CAGGTTCCTC GACTGGAAG CGGGCAGTGA
6901 GGGCAUGCA ATTAATGTA GTTAGCTCAC TCATTAGGCCA CCCCAGGCTT
6951 TACACTTAT GCTTCCGGCT CGTATGTTG GTGGAATTGT GAGGGATAAA
7001 CAATTCAACA CAGGAAACAG CTATGAC

FIG 5B CONT.

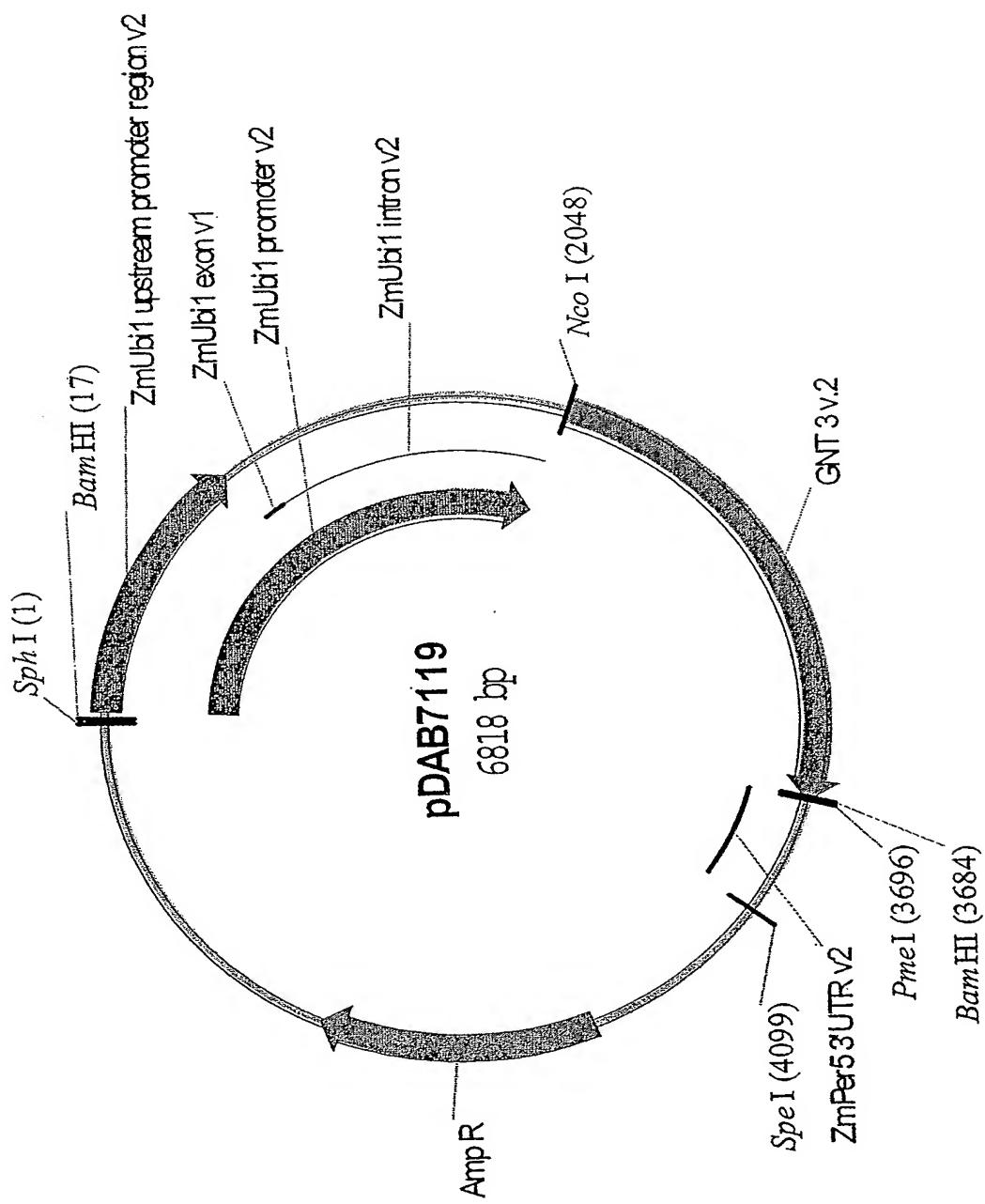


FIG. 6A

pDAB7119 6818bp

Sequence	Feature
1-38 Linker	CCTGCAGATCCCCGGGGATCCTCTAGAGTCGACCTGCA
39-2028	Maize Ubiquitin 1 promoter
2029-2047 Linker	GGGTACCCCCGGGGTCGAC
2048-3695	GNT III v.2 plus TAGTTT
2462	C to replace G as reported in original sequence
3696-3699 PmeI	AAAC
3700-4064	Maize Per 5 3'UTR
4065-6818	pUC19 backbone
4259-4264	TGCGCA FspI
4720-5580	Ampicillin Resistance gene
5282-5287	TGCGCA FspI

BamHI

~~~~~

1 CCTGCAGATC CCCGGGGATC CTCTAGAGTC GACCTGCAGT  
 GCAGCGTGAC CCGGTCGTGC CCCTCTCTAG AGATAATGAG  
 CATTGCATGT CTAAGTTATA  
 101 AAAAATTACC ACATATTTTT TTTGTCACAC TTGTTTGAAG  
 TGCAGTTTAT CTATCTTTAT ACATATATTT AAACTTTAAT  
 CTACGAATAA TATAATCTAT  
 201 AGTACTACAA TAATATCAGT GTTTAGAGA ATCATATAAA  
 TGAACAGTTA GACATGGTCT AAAGGACAAT TGAGTATTCTT  
 GACAACAGGA CTCTACAGTT  
 301 TTATCTTTTT AGTGTGCATG TGTTCTCCTT TTTTTTGCA  
 AATAGCTTCA CCTATATAAT ACTTCATCCA TTTTATTAGT  
 ACATCCATT AGGGTTTAGG  
 401 GTTAATGGTT TTTATAGACT AATTTTTTA GTACATCTAT  
 TTTATTCTAT TTTAGCCTCT AAATTAAGAA AACTAAAAT  
 CTATTTAGT TTTTTTATTT  
 501 AATAATTTAG ATATAAAATA GAATAAAATA AAGTGACTAA  
 AAATTAAACA AATACCCTTT AAGAAATTAA AAAAACTAAG  
 GAAACATTTT TCTTGTTCG  
 601 AGTAGATAAT GCCAGCCTGT TAAACGCCGT CGACGAGTCT  
 AACGGACACC AACCAGCGAA CCAGCAGCGT CGCGTCGGGC  
 CAAGCGAAGC AGACGGCAGC  
 701 GCATCTCTGT CGCTGCCCTCT GGACCCCTCT CGAGAGTTCC  
 GCTCCACCGT TGGACTTGCT CCGCTGTCGG CATCCAGAAA  
 TTGCGTGGCG GAGCGGCAGA  
 801 CGTGAGCCGG CACGGCAGGC GGCCTCCTCC TCCTCTCACG  
 GCACGGCAGC TACGGGGAT TCCTTCCC A CCGCTCCTTC  
 GCTTTCCCTT CCTCGCCCCGC

**FIG 6B**

901 CGTAATAAAT AGACACCCCC TCCACACCCT CTTTCCCCAA  
 CCTCGTGTG TTCGGAGCGC ACACACACAC AACCAAGATCT  
 CCCCCAAATC CACCCGTGCG  
 1001 CACCTCCGCT TCAAGGTACG CCGCTCGTCC TCCCCCCCCC  
 CCCCTCTCTA CCTTCTCTAG ATCGGCCTTC CGGTCCATGC  
 ATGGTTAGGG CCCGGTAGTT  
 1101 CTACTTCTGT TCATGTTGT GTTAGATCCG TGTTTGTGTT  
 AGATCCGTGC TGCTAGCGTT CGTACACCGA TGCGACCTGT  
 ACGTCAGACA CGTTCTGATT  
 1201 GCTAACTTGC CAGTGTTCCT CTTGGGGAA TCCTGGGATG  
 GCTCTAGCCG TTCCGCAGAC GGGATCGATT TCATGATTTC  
 TTTGTTTCG TTGCATAGGG  
 1301 TTTGGTTTGC CCTTTTCCTT TATTTCATAA TATGCCGTGC  
 ACTTGTTTGT CGGGTCATCT TTTCATGCTT TTTTTGTCT  
 TGGTTGTGAT GATGTGGTCT  
 1401 GGTTGGCGG TCGTTCTAGA TCGGAGTAGA ATTCTGTTTC  
 AAACTACCTG GTGGATTTAT TAATTTGGA TCTGTATGTG  
 TGTGCCATAC ATATTCA TAG  
 1501 TTACGAATTG AAGATGATGG ATGGAAATAT CGATCTAGGA  
 TAGGTATAACA TGTTGATGCG GGTTTTACTG ATGCATATAC  
 AGAGATGCTT TTTGTTCGCT  
 1601 TGGTTGTGAT GATGTGGTGT GGTTGGCGG TCGTTCATTC  
 GTTCTAGATC GGAGTAGAAT ACTGTTCAA ACTACCTGGT  
 GTATTATTA ATTTTGGAAC  
 1701 TGTATGTGTG TGTACATACAT CTTCATAGTT ACGAGTTAA  
 GATGGATGGA AATATCGATC TAGGATAGGT ATACATGTTG  
 ATGTGGGTTT TACTGATGCA  
 1801 TATACATGAT GGCATATGCA GCATCTATTTC ATATGCTCTA  
 ACCTTGAGTA CCTATCTATT ATAATAAACAA AGTATGTTTT  
 ATAATTATTG TGATCTTGAT  
 1901 ATACTTGGAT GATGGCATAT GCAGCAGCTA TATGTGGATT  
 TTTTAGCCC TGCCCTCATA CGCTATTTAT TTGCTTGGTA  
 CTGTTCTTT TGTCGATGCT

NcoI

FseI

~~~~~

~~~~~

2001 CACCCTGTG TTTGGTGTAA CTTCTGCAGG GTACCCCCGG  
 GGTGACCAT GGTGATGAGA CGCTACAAGC TCTTCTCAT  
 GTTCTGTATG GCCGGCTGT  
 2101 GCCTCATCTC CTTCCCTGCAC TTCTTCAAGA CCCTGTCCTA  
 TGTCACCTTC CCCCCGAGAAC TGGCCTCCCT CAGCCCTAAC  
 CTGGTGTCCA GCTTTTCTG  
 2201 GAACAATGCC CCGGTCACGC CCCAGGCCAG CCCCGAGCCA  
 GGAGGCCCTG ACCTGCTGCG TACCCCACTC TACTCCCACT  
 CGCCCCTGCT GCAGCCGCTG

**SacI****FspI**

~~~~~

~~~~~

2301 CCGCCCAGCA AGGCGGCCGA GGAGCTCCAC CGGGTGGACT  
       TGGTGCTGCC CGAGGACACC ACCGAGTATT TCGTGCGCAC  
       CAAGGCCGGC GGCGTCTGCT  
 2401 TCAAACCCGG CACCAAGATG CTGGAGAGGC CCCCCCCCGGG  
       ACGGCCGGAG GAGAAGCCTG AGGGGGCCAA CGGCTCCTCG  
       GCCCGCGGC CACCCCGGTA  
 2501 CCTCCTGAGC GCCCGGGAGC GCACGGGGGG CCGAGGCGCC  
       CGGCGCAAGT GGGTGGAGTG CGTGTGCCTG CCCGGCTGGC  
       ACGGACCCAG CTGCGGCGTG  
 2601 CCCACTGTGG TGCA GTACTC CAACCTGCC ACCAAGGAGC  
       GGCTGGTGCC CAGGGAGGTG CCGCGCCGCG TCATCAACGC  
       CATCAACGTC AACACAGAGT

**NotI**

~~~~~

2701 TCGACCTGCT GGACGTGCGC TTCCACGAGC TGGGCGACGT
 GGTGGACGCC TTTGTGGTGT GCGAGTCAA CTTCACGGCT
 TATGGGGAGC CGCGGCCGCT
 2801 CAAGTTCCGG GAGATGCTGA CCAATGGCAC CTTCGAGTAC
 ATCCGCCACA AGGTGCTCTA TGTCTTCCTG GACCACTTCC
 CGCCCGGC GGCGCAGGAC

FspI**FspI**

~~~~~

~~~~~

2901 GGCTGGATCG CCGACGACTA CCTGCGCACC TTCCTCACCC
 AGGACGGCGT CTCGCGGCTG CGAACCTGC GGCCCGACGA
 CGTCTTCATC ATTGACGATG

FspI

~~~~~

3001 CGGACGAGAT CCCGGCCCGT GACGGCGTCC TTTTCCTCAA  
       GCTCTACGAT GGCTGGACCG AGCCCTTCGC CTTCCACATG  
       CGCAAGTCGC TCTACGGCTT  
 3101 CTTCTGGAAG CAGCCGGCA CCCTGGAGGT GGTGTCAGGC  
       TGCACGGTGG ACATGCTGCA GGCAGTGTAT GGGCTGGACG  
       GCATCCGCCT GCGCCGCGC  
 3201 CAGTACTACA CCATGCCAA CTTCAGACAG TATGAGAACCC  
       GCACCGGCCA CATCCTGGTG CAGTGGTCGC TGGGCAGCCC  
       CCTGCACTTC GCCGGCTGGC

FIG 6B CONT.

3301 ACTGCTCCTG GTGCTTCACG CCCGAGGGCA TCTACTTCAA  
           GCTCGTGTCC GCCCAGAATG GCGACTTCCC ACGCTGGGGT  
           GAATACGAGG ACAAGCGGGA  
 3401 CCTGAACTAC ATCCCGGGCC TGATCCGCAC CGGGGGCTGG  
           TTCGACGGCA CGCAGCAGGA GTACCCGCCT GCAGACCCCA  
           GCGAGCACAT GTATGCGCCC  
 3501 AAGTACCTGC TGAAGAACTA CGACCGGTT CACTACCTGC  
           TGGACAACCC CTACCAGGAG CCCAGGAGCA CGGC GGCGGG  
           CGGGTGGCGC CACAGGGGTC

BamHI                  PmeI

~~~~~ ~~~~~~

3601 CCGAGGGAAAG GCCGCCCGCC CGGGGCAAAC TGGACGGAGGC
 GGAAGTCGAA CAAAAACTCA TCTCAGAAGA GGATCTGAAT
 TAGGATCCTA GGTTTAAACT
 3701 GAGGGCACTG AAGTCGCTTG ATGTGCTGAA TTGTTTGTGA
 TGTTGGTGGC GTATTTGTT TAAATAAGTA AGCATGGCTG
 TGATTTTATC ATATGATCGA
 3801 TCTTTGGGGT TTTATTTAAC ACATTGTAAA ATGTGTATCT
 ATTAATAACT CAATGTATAA GATGTGTTCA TTCTTCGGTT
 GCCATAGATC TGCTTATTG
 3901 ACCTGTGATG TTTGACTCC AAAAACCAAA ATCACAACTC
 AATAAAACTCA TGGAATATGT CCACCTGTTT CTTGAAGAGT
 TCATCTACCA TTCCAGTTGG

FseI

~~~~~

4001 CATTATTCAG TGTTGCAGCG GCGCTGTGCT TTGTAACATA  
           ACAATTGTTCA CGGGCATATA TCCACGGCCG GCCTAGCTAG  
           CCACGGTGGC CAGATCCACT

NotI

~~~~~

4101 AGTTCTAGAG CGGCCGCTTA ATTCACTGGC CGTCGTTTA
 CAACGTCGTG ACTGGGAAAA CCCTGGCGTT ACCCAACTTA
 ATCGCCTTGC AGCACATCCC

FspI

~~~~~

4201 CCTTCGCCA GCTGGCGTAA TAGCGAAGAG GCCCGCACCG  
           ATCGCCCTTC CCAACAGTTG CGCAGCCTGA ATGGCGAATG  
           GCGCCTGATG CGGTATTTTC  
 4301 TCCTTACGCA TCTGTGCGGT ATTTCACACC GCATATGGTG  
           CACTCTCAGT ACAATCTGCT CTGATGCCGC ATAGTTAAGC  
           CAGCCCCGAC ACCCGCCAAC

4401 ACCCGCTGAC GCGCCCTGAC GGGCTTGTCT GCTCCCGGCA  
 TCCGCTTACA GACAAGCTGT GACCGTCTCC GGGAGCTGCA  
 TGTGTCAGAG GTTTCACCG  
 4501 TCATCACCGA AACGCGCGAG ACGAAAGGGC CTCGTGATAC  
 GCCTATTTC ATAGGTTAAC GTCATGATAA TAATGGTTTC  
 TTAGACGTCA GGTGGCACTT  
 4601 TTCGGGGAAA TGTGCGCGGA ACCCCTATTG GTTTATTTC  
 CTAATACAT TCAAATATGT ATCCGCTCAT GAGACAATAA  
 CCCTGATAAA TGCTTCATAA  
 4701 ATATTGAAAA AGGAAGAGTA TGAGTATTCA ACATTCCGT  
 GTCGCCCTTA TTCCCTTTTG TGCAGCATTG TGCCTTCCTG  
 TTTTGCTCA CCCAGAAACG  
 4801 CTGGTGAAG TAAAAGATGC TGAAGATCAG TTGGGTGCAC  
 GAGTGGGTTA CATCGAACTG GATCTCAACA GCGGTAAGAT  
 CCTTGAGAGT TTTCGCCCCG  
 4901 AAGAACGTTT TCCAATGATG AGCACTTTA AAGTTCTGCT  
 ATGTGGCGCG GTATTATCCC GTATTGACGC CGGGCAAGAG  
 CAACTCGGTC GCCGCATAACA  
 5001 CTATTCTCAG AATGACTTGG TTGAGTACTC ACCAGTCACA  
 GAAAAGCATC TTACGGATGG CATGACAGTA AGAGAATTAT  
 GCAGTGCTGC CATAACCATG  
 5101 AGTGATAACA CTGCGGCCAA CTTACTTCTG ACAACGATCG  
 GAGGACCGAA GGAGCTAACCG GCTTTTTGC ACAACATGGG  
 GGATCATGTA ACTCGCCTTG

## FspI

~~~~~

5201 ATCGTTGGGA ACCGGAGCTG AATGAAGCCA TACCAAACGA
 CGAGCGTGAC ACCACGATGC CTGTAGCAAT GGCAACAAACG
 TTGCGCAAAC TATTAACCTGG
 5301 CGAACTACTT ACTCTAGCTT CCCGGCAACA ATTAATAGAC
 TGGATGGAGG CGGATAAAGT TGCAGGACCA CTTCTGCGCT
 CGGCCCTTCC GGCTGGCTGG
 5401 TTTATTGCTG ATAAATCTGG AGCCGGTGAG CGTGGGTCTC
 GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC
 CCGTATCGTA GTTATCTACA
 5501 CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA
 GATCGCTGAG ATAGGTGCCT CACTGATTAA GCATTGGTAA
 CTGTCAGACC AAGTTACTC
 5601 ATATATACTT TAGATTGATT TAAAACCTCA TTTTTAATT
 AAAAGGATCT AGGTGAAGAT CCTTTTGAT AATCTCATGA
 CCAAAATCCC TTAACGTGAG
 5701 TTTTCGTTCC ACTGAGCGTC AGACCCCGTA GAAAAGATCA
 AAGGATCTTC TTGAGATCCT TTTTTCTGC GCGTAATCTG
 CTGCTTGCAA ACAAAAAAAC

5801 CACCGCTACC AGCGGTGGTT TGTTGCCGG ATCAAGAGCT
 ACCAACTCTT TTTCCGAAGG TAACTGGCTT CAGCAGAGCG
 CAGATACCAA ATACTGTCCT
 5901 TCTAGTGTAG CCGTAGTTAG GCCACCACCT CAAGAACTCT
 GTAGCACCGC CTACATACCT CGCTCTGCTA ATCCTGTTAC
 CAGTGGCTGC TGCCAGTGGC
 6001 GATAAGTCGT GTCTTACCGG GTTGGACTCA AGACGATAGT
 TACCCGATAA GGCGCAGCGG TCGGGCTGAA CGGGGGGTTTC
 GTGCACACAG CCCAGCTTGG
 6101 AGCGAACGAC CTACACCGAA CTGAGATACC TACAGCGTGA
 GCATTGAGAA AGCGCCACGC TTCCCGAAGG GAGAAAGGCG
 GACAGGTATC CGGTAAGCGG
 6201 CAGGGTCGGA ACAGGAGAGC GCACGAGGGGA GCTTCCAGGG
 GGAAACGCCT GGTATCTTTA TAGTCCTGTC GGGTTTCGCC
 ACCTCTGACT TGAGCGTCGA
 6301 TTTTGTTGAT GCTCGTCAGG GGGGCGGAGC CTATGGAAAA
 ACGCCAGCAA CGCGGCCTTT TTACGGTTCC TGGCCTTTG
 CTGGCCTTTT GCTCACATGT
 6401 TCTTTCCTGC GTTATCCCCT GATTCTGTGG ATAACCGTAT
 TACCGCCTTT GAGTGAGCTG ATACCGCTCG CCGCAGCCGA
 ACGACCGAGC GCAGCGAGTC
 6501 AGTGAGCGAG GAAGCGGAAG AGCGCCCAAT ACGCAAACCG
 CCTCTCCCCG CGCGTTGGCC GATTCATTA TGCAAGCTGGC
 ACGACAGGTT TCCCGACTGG
 6601 AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC
 TCACTCATTA GGCACCCCCAG GCTTACACT TTATGCTTCC
 GGCTCGTATG TTGTGTGGAA

Not I

~~~~~

6701 TTGTGAGCGG ATAACAATT CACACAGGAA ACAGCTATGA  
 CCATGATTAC GCCAAGCTAG CGGCCGCATT CCCGGGAAGC  
 TAGGCCACCG TGGCCCGCCT

## HindIII

~~~~~

6801 GCAGGGGAAG CTTGCATG

FIG 6B CONT.

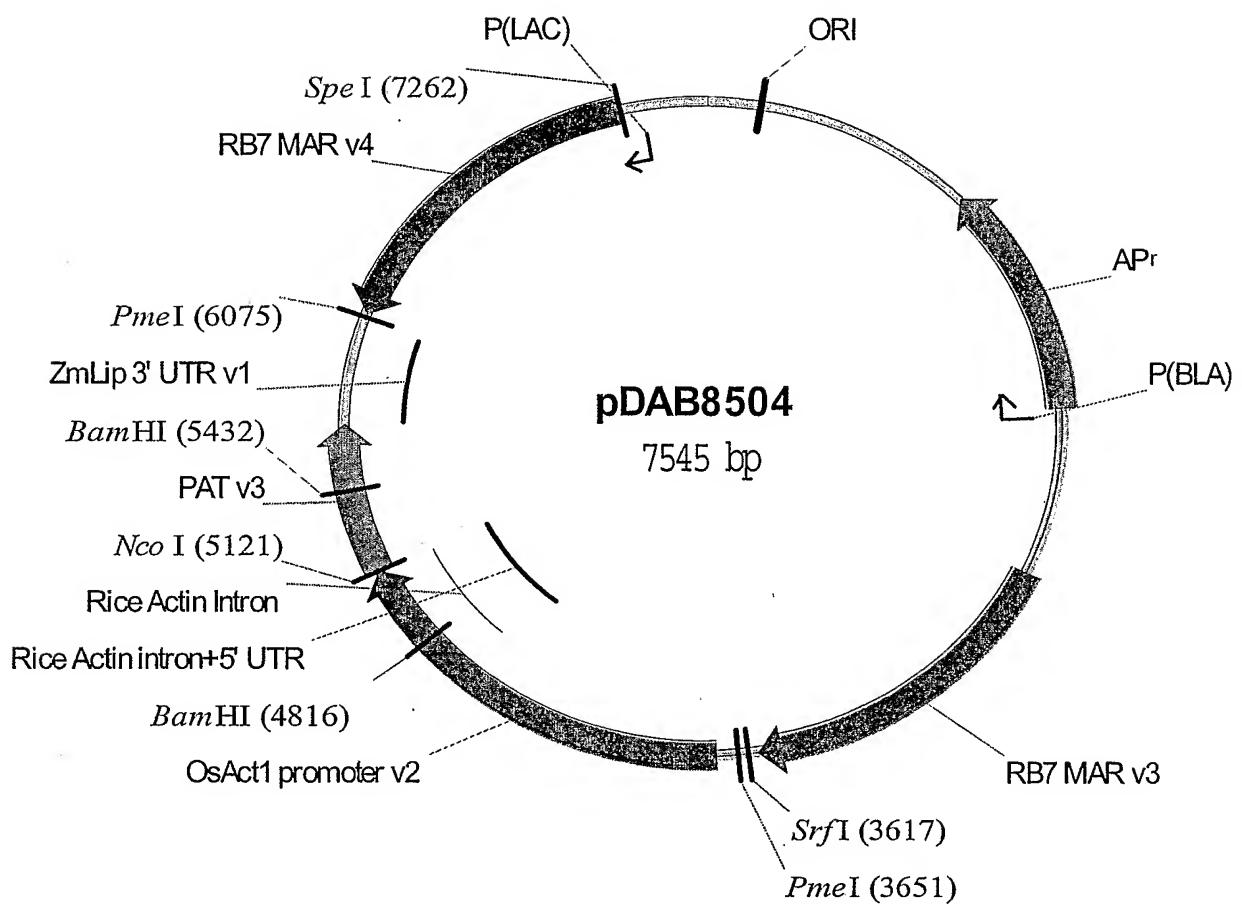


FIG. 7A

Plasmid pDAB8504 (7545 bp)

Description: Cloning vector with Rb7 MARs (inverted orientation) flanking a multiple cloning site and the rice actin/PAT/lipase selectable marker cassette

| Nucleotide start | Nucleotide end | Sequence Feature |
|------------------|----------------|--|
| 1 | 1166 | Tobacco Rb7 MARs |
| 1167 | 1304 | Linker Sequence with multiple cloning site:
TGGCCACCGCTTAATTAAGGCGCGCCATGCCCGGGCAAGCG
GCGCCTAATTAAATTAAATGTTAAACTAGGAAATCCAA
GCTTGGGCTGCAGGTCAATCCCATTGCTTTGAAGCAGCTC
AACATTGATCTCTT |
| 1305 | 2701 | Rice actin promoter with intron |
| 2235 | 2696 | Rice actin intron |
| 2702 | 2703 | CC |
| 2704 | 3258 | PAT gene (phosphinothricin acyl transferase) |
| 3259 | 3272 | Linker sequence: GGTACCCTGAGCTC |
| 3273 | 3629 | Maize lipase 3' UTR |
| 3630 | 3670 | Linker sequence:
GAATTCATATTCCTCCTGCAGGGTTAAACTGCCGTGGC |
| 3671 | 4836 | Tobacco Rb7 MAR (complementary) |
| 4837 | 4857 | Linker sequence: CGGCCCACTAGTCACCGGTGT |
| 4858 | 5103 | Puc19 |
| 5104 | 5130 | Linker sequence:
GCGCACGCTGCGCACGCTGCGCACGCT |
| 5130 | 7523 | Puc19 |
| 7524 | 7545 | Linker sequence: ACACCGGTGTGATCATGGGCCG |

Sequence

1 CGATTAAAAA TCTCAATTAT ATTTGGTCTA ATTTAGTTG GTATTGAGTA
 51 AAACAAATTC GAACCAAACC AAAATATAAA TATATAGTTT TTATATATAT
 101 GCCTTAAAGA CTTTTATAG AATTTCTTT AAAAAATATC TAGAAATATT
 151 TGC GACT CTT CTGGCATGTA ATATTCGTT AAATATGAAG TGCTCCATTT
 201 TTATTAAC TT AAAATAATTG GTTGTACGAT CACTTCTTA TCAAGTGT
 251 CTAAAATGCG TCAATCTCTT TGTCTTCCA TATT CATATG TCAAAACCTA
 301 TCAAAATTCT TATATATCTT TTTCGAATTG GAAGTGAAAT TTCGATAATT
 351 TAAAATTAAA TAGAACATAT CATTATTAG GTATCATATT GATTTTTATA
 401 CTTAATTACT AAAATTGGTT AAC TTTGAAA GTGTACATCA ACGAAAATT
 451 AGTCAAACGA CTAAAATAAA TAAATATCAT GTGTTATTAA GAAAATTCTC
 501 CTATAAGAAT ATTTTAATAG ATCATATGTT TGTA AAAAAAA ATTAAATT
 551 ACTAACACAT ATATTTACTT ATCAAAATT TGACAAAGTA AGATTA
 601 AATATTCA TC AACAAAAAA AAAACCAGAA AATGCTGAAA ACCCGGCAA
 651 ACCGAACCAA TCCAAACCGA TATAGTTGGT TTGGTTTGAT TTTGATATAA
 701 ACCGAACCAA CTCGGTCCAT TTGCACCCCT AAC CATAATA GCTTTAATAT
 751 TTCAAGATAT TATTAAGTTA AC GTTGTCAA TATCCTGGAA ATTTGCAA
 801 ATGAATCAAG CCTATATGGC TGTAATATGA ATTTAAAAGC AGCTCGATGT
 851 GGTGGTAATA TGTAATTTAC TTGATTCTAA AAAAATATCC CAAGTATTAA

901 TAATTTCTGC TAGGAAGAAG GTTAGCTACG ATTTACAGCA AAGCCAGAAT
 951 ACAATGAACC ATAAAGTGT TGAAGCTCGA AATATAACGAA GGAACAAATA
 1001 TTTTTAAAAA AATACGCAAT GACTTGGAAC AAAAGAAAAGT GATATATT
 1051 TTGTTCTTAA ACAAGCATCC CCTCTAAAGA ATGGCAGTTT TCCTTGCAT
 1101 GTAACATTAA TGCTCCCTTC GTTACAAAAA TTTGGACTA CTATTGGAA
 1151 CTTCTTCTGA AAATAGTGGC CACCGCTTAA TTAAGGCGCG CCATGCCGG
 1201 GCAAGCGGCC GCTTAATTAA ATTTAAATGT TAAACTAGG AAATCCAAGC
 1251 TTGGGCTGCA GGTCAATCCC ATTGCTTTG AAGCAGCTCA ACATTGATCT
 1301 CTTCTCGAG GTCATTCTATA TGCTTGAGAA GAGAGTCGGG ATAGTCCAAA
 1351 ATAAAACAAA GGTAAGATTA CCTGGTCAAAG AGTAAAACA TCAGTTAAA
 1401 GGTGGTATAA AGTAAAATAT CGGTAATAAA AGGTGGCCA AAGTGAATT
 1451 TACTCTTTTC TACTATTATA AAAATTGAGG ATGTTTTGT CGGTACTTTG
 1501 ATACGTCATT TTTGTATGAA TTGGTTTTA AGTTTATTGCG CTTTTGGAAA
 1551 TGCATATCTG TATTTGAGTC GGGTTTTAAG TTCGTTTGCT TTTGTAATA
 1601 CAGAGGGATT TGTATAAGAA ATATCTTAA AAAAACCCAT ATGCTAATT
 1651 GACATAATT TTAGAGAAAAA TATATATTCA GGCAGATTCT CACAATGAAC
 1701 AATAATAAGA TTAAAATAGC TTTCCCCCGT TGAGCGCAT GGGTATT
 1751 TCTAGTAAAAA ATAAAAGATA AACTTAGACT CAAAACATT ACACAAACAA
 1801 CCCCTAAAGT TCCTAAAGCC CAAAGTGTCA TCCACGATCC ATAGCAAGCC
 1851 CAGCCCAACC CAACCCAAACC CAACCCACCC CAGTCCAGCC AACTGGACAA
 1901 TAGTCTCCAC ACCCCCCCAC TATCACCCTG AGTTGTCCGC ACGCACCGCA
 1951 CGTCTCGCAG CCAAAAAAAA AAAAGAAAG AAAAGAAAAA AAAAGAAAAA
 2001 ACAGCAGGTG GGTCCGGGTC GTGGGGGCCG GAAACGCGAG GAGGATCGCG
 2051 AGCCAGCGAC GAGGCCGGCC CTCCCTCCGC TTCCAAAGAA ACGCCCCCA
 2101 TCGCCACTAT ATACATACCC CCCCCCTCTCC TCCCCTCCCC CCAACCTAC
 2151 CACCACCACC ACCACCACCT CCACCTCCTC CCCCCCTCGCT GCCGGACGAC
 2201 GCCTCCCCCCC TCCCCCTCCG CCGCCGCCGC GCCGGTAACC ACCCCGCC
 2251 TCTCCTCTTT CTTTCTCCGT TTTTTTTTC CGTCTCGGTC TCGATCTTTG
 2301 GCCTGGTAG TTTGGGTGGG CGAGAGGCGG CTTCTGCGC GCCCAGATCG
 2351 GTGCGCGGGA GGGCGGGGAT CTCGCGGCTG GGGCTCTCGC CGGCGTGGAT
 2401 CCGGCCCCGA TCTCGCGGGG AATGGGGCTC TCGGATGTAG ATCTGCGATC
 2451 CGCCGTTGTT GGGGGAGATG ATGGGGGTT TAAAATTCC GCCATGCTAA
 2501 ACAAGATCAG GAAGAGGGGA AAAGGGCACT ATGGTTATA TTTTTATATA
 2551 TTTCTGCTGC TTCTGTCAGGC TTAGATGTGC TAGATCTTTC TTTCTTCTTT
 2601 TTGTGGGTAG AATTGAATC CCTCAGCATT GTTCATCGGT AGTTTTCTT
 2651 TTCAATGATTT GTGACAAATG CAGCCTCGTG CGGAGCTTT TTGTAGGTAG
 2701 ACCATGGCTT CTCCGGAGAG GAGACCAGT GAGATTAGGC CAGCTACAGC
 2751 AGCTGATATG GCCGCGGTT GTGATATCGT TAACCATTAC ATTGAGACGT
 2801 CTACAGTGA A CTTTAGGACA GAGCCACAAA CACCACAAGA GTGGATTGAT
 2851 GATCTAGAGA GGTTGCAAGA TAGATACCT TGTTGGTTG CTGAGGTTGA
 2901 GGGTGTGTTG GCTGGTATTG CTTACGCTGG GCCCTGGAAG GCTAGGAACG
 2951 CTTACGATTG GACAGTTGAG AGTACTGTT AGTGTCAAC TAGGCATCAA
 3001 AGGTTGGGCC TAGGATCCAC ATTGTACACA CATTGCTTA AGTCTATGGA
 3051 GGCAGCAAGGT TTTAAGTCTG TGTTGCTGT TATAGGCCTT CCAAACGATC
 3101 CATCTGTTAG GTTGCATGAG GCTTTGGGAT ACACAGCCCG GGGTACATTG
 3151 CGCGCAGCTG GATACAAGCA TGGTGGATGG CATGATGTTG GTTTTGGCA
 3201 AAGGGATTT GAGTTGCCAG CTCCTCCAAG GCCAGTTAGG CCAGTTACCC
 3251 AGATCTGAGG TACCCCTGAGC TCGGTGCGCAG CGTGTGCGTG TCCGTCGTAC
 3301 GTTCTGGCCG GCCGGGCCTT GGGCGCGCGA TCAGAAGCGT TGCCTGGCG
 3351 TGTGTGTGCT TCTGGTTGCT TTTAATTAA CCAAGTTGT TTCAAGGTGG
 3401 ATCGCGTGGT CAAGGCCGT GTGCTTAAA GACCCACCGG CACTGGCAGT
 3451 GAGTGTGCT GCTTGGTAG GCTTGGTAC GTATGGCCTT TATTGCTTC

3501 TGGATGTTGT GTACTACTTG GGTTTGTGA ATTATTATGA GCAGTTGCGT
 3551 ATTGTAATTC AGCTGGGCTA CCTGGACATT GTTATGTATT AATAAATGCT
 3601 TTGCTTTCTT CTAAAGATCT TTAAGTGTG AATTCAATT TCCTCCTGCA
 3651 GGGTTTAAAC TTGCCGTGGC CTATTTCAAG AAGAAGTTCC CAATAGTAGT
 3701 CCAAAATTT TGTAACGAAG GGAGCATAAT AGTTACATGC AAAGGAAAAC
 3751 TGCCATTCTT TAGAGGGAT GCTTGTAA GAACAAAAAA TATATCACTT
 3801 TCTTTGTTCAAGTCATTG CGTATTTTT TAAAATATT TGTTCCTTCG
 3851 TATATTCGA GCTTCAATCA CTTTATGGTT CTTGTATTG TGGCTTTGCT
 3901 GTAAATCGTA GCTAACCTTC TTCCTAGCAG AAATTATTAA TACTTGGGAT
 3951 ATTTTTTAG AATCAAGTAA ATTACATATT ACCACCACAT CGAGCTGCTT
 4001 TTAAATTCAAT ATTACAGCCA TATAGGCTTG ATTCATTTG CAAAATTCC
 4051 AGGATATTGA CAACGTTAAC TTAATAATAT CTGAAATAT TAAAGCTATT
 4101 ATGATTAGGG GTGCAAATGG ACCGAGTTGG TTCGGTTAT ATCAAAATCA
 4151 AACCAAACCA ACTATATCGG TTTGGATTGG TTCGGTTTG CCGGGTTTC
 4201 AGCATTTCGTTGGTTTT TTGTTAGATG AATATTATT TAATCTTACT
 4251 TTGTCAAATT TTTGATAAGT AAATATATGT GTTAGTAAAA ATTAATTTTT
 4301 TTTACAAACA TATGATCTAT TAAAATATT TTATAGGAGA ATTTCTTAA
 4351 TAACACATGA TATTATTAA TTTAGTCGT TTGACTAATT TTTCGTTGAT
 4401 GTACACTTTC AAAGTTAACCA AAATTAGTA ATTAAGTATA AAAATCAATA
 4451 TGATACCTAA ATAATGATAT GTTCTATTAA ATTTAAATT ATCGAAATTT
 4501 CACTCAAAT TCGAAAAAGA TATATAAGAA TTTTGATAGA TTTTGACATA
 4551 TGAATATGGA AGAACAAAGA GATTGACGCA TTTTAGTAAC ACTTGATAAG
 4601 AAAGTGATCG TACAACCAAT TATTAAAGT TAATAAAAT GGAGCACTTC
 4651 ATATTAACG AAATATTACA TGCCAGAAGA GTCGCAAATA TTTCTAGATA
 4701 TTTTTAAAG AAAATTCTAT AAAAGTCTT AAAGGCATAT ATATAAAAC
 4751 TATATATTAA TATTTGGTT TGGTCGAAT TTGTTTACT CAATACAAA
 4801 CTAAATTAGA CCAAATATAA TTGGGATTT TAATCGCGGC CCACTAGTC
 4851 CCGGTGTAAT TTGGCGTAAT CATGGTCATA GCTGTTTCCT GTGTGAAATT
 4901 GTTATCCGCT CACAATTCCA CACAACATAC GAGCCGGAAG CATAAAGTGT
 4951 AAAGCCTGGG GTGCCTAATG AGTGAGCTAA CTCACATTAA TTGCGTTGCG
 5001 CTCACTGCC CTTTCCAGT CGGGAAACCT GTCGTGCCAG CTGCATTAAT
 5051 GAATCGGCCA ACGCGGGGG AGAGCGGGTT TCGGTATTGG GCGCTTTCC
 5101 GCTGCGCACG CTGCGCACGC TGCGCACGCT TCCTCGCTCA CTGACTCGCT
 5151 GCGCTCGGTC GTTCGGCTGC GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG
 5201 TAATACGGTT ATCCACAGAA TCAGGGATA ACGCAGGAAA GAACATGTGA
 5251 GCAAAAGGCC AGCAAAAGGC CAGGAACCGT AAAAAGGCCG CGTTGCTGGC
 5301 GTTTTCCAT AGGCTCGGCC CCCCTGACGA GCATCACAAA AATCGACGCT
 5351 CAAGTCAGAG GTGGCGAAAC CCGACAGGAC TATAAAGATA CCAGGCGTTT
 5401 CCCCCTGGAA GCTCCCTCGT GCGCTCTCCT GTTCCGACCC TGCCGCTTAC
 5451 CGGATACCTG TCCGCCCTTC TCCCTTCGGG AAGCGTGGCG CTTTCTCATA
 5501 GCTCACGCTG TAGGTATCTC AGTCGGTGT AGTCGTTCG CTCCAAGCTG
 5551 GGCTGTGTGC ACGAACCCCC CGTTCAGCCC GACCGCTGCG CCTTATCCGG
 5601 TAACTATCGT CTTGAGTCCA ACCCGGTAAG ACACGACTTA TCGCCACTGG
 5651 CAGCAGCCAC TGGTAACAGG ATTAGCAGAG CGAGGTATGT AGGCAGGTGCT
 5701 ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA GAAGGACAGT
 5751 ATTTGGTATC TGCGCTCTGC TGAAGCCAGT TACCTTCGGA AAAAGAGTTG
 5801 GTAGCTCTTG ATCCGGAAA CAAACCACCG CTGGTAGCGG TGGTTTTTT
 5851 GTTTGCAAGC AGCAGATTAC GCGCAGAAAA AAAGGATCTC AAGAAGATCC
 5901 TTTGATCTT TCTACGGGTT CTGACGCTCA GTGGAACGAA AACTCACGTT
 5951 AAGGGATTTT GGTCAATGAGA TTATCAAAAA GGATCTTCAC CTAGATCCTT
 6001 TTAAATTAAA AATGAAGTTT TAAATCAATC TAAAGTATAT ATGAGTAAAC
 6051 TTGGTCTGAC AGTTACCAAT GCTTAATCAG TGAGGCACCT ATCTCAGCGA

6101 TCTGTCATT TCGTTCATCC ATAGTTGCCCT GACTCCCCGT CGTAGATA
6151 ACTACGATAC GGGAGGGCTT ACCATCTGGC CCCAGTGCTG CAATGATACC
6201 GCGAGACCCA CGCTCACCGG CTCCAGATT ATCAGCAATA AACCAGCCAG
6251 CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG CAACTTTATC CGCCTCCATC
6301 CAGTCTATTA ATTGTTGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA
6351 TAGTTGCGC AACGTTGTTG CCATTGCTAC AGGCATCGTG GTGTCACGCT
6401 CGTCGTTGG TATGGCTTCA TTCAGCTCCG GTTCCAACG ATCAAGGCGA
6451 GTTACATGAT CCCCCATGTT GTGAAAAAA CGGGTTAGCT CCTTCGGTCC
6501 TCCGATCGTT GTCAGAAGTA AGTTGCCGC AGTGTATCA CTCATGGTTA
6551 TGGCAGCACT GCATAATTCT CTTACTGTCA TGCCATCCGT AAGATGCTTT
6601 TCTGTGACTG GTGAGTACTC AACCAAGTCA TTCTGAGAAT AGTGTATGCG
6651 GCGACCGAGT TGCTCTGCC CGCGTCAAT ACAGGGATAAT ACCGCGCCAC
6701 ATAGCAGAAC TTTAAAAGTG CTCATCATTG GAAAACGTT TCAGGGCGA
6751 AAACCTCTCAA GGATCTTACC GCTGTTGAGA TCCAGTTCGA TGTAACCCAC
6801 TCGTGCACCC AACTGATCTT CAGCATCTT TACTTCACC AGCGTTCTG
6851 GGTGAGCAAA AACAGGAAGG CAAAATGCCG CAAAAAAGGG AATAAGGGCG
6901 ACACGGAAAT GTTGAATACT CATACTCTTC CTTTTCAAT ATTATTGAAG
6951 CATTATCAG GGTTATTGTC TCATGAGCGG ATACATATT GAATGTATTT
7001 AGAAAAATAA ACAAAATAGGG GTTCCCGCACA CATTTCGGG AAAAGTGCCA
7051 CCTGACGTCT AAGAAACCAT TATTATCATG ACATTAACCT ATAAAAATAG
7101 GCGTATCAG AGGCCCTTTC GTCTCGCGCG TTTCGGTGAT GACGGTGAAA
7151 ACCTCTGACA CATGCAGCTC CGGGAGACGG TCACAGCTT TCTGTAAGCG
7201 GATGCCGGGA GCAGACAAGC CCGTCAGGGC GCGTCAGCGG GTGTTGGCGG
7251 GTGTCGGGGC TGGCTTAACG ATGCGGCATC AGAGCAGATT GTACTGAGAG
7301 TGCACCATAT GCGGTGTGAA ATACCGCACA GATGCGTAAG GAGAAAATAC
7351 CGCATCAGGC GCCATTGCC ATTCAAGGCTG CGCAACTGTT GGGAAAGGGCG
7401 ATCGGTGCGG GCCTCTTCGC TATTACGCCA GCTGGCGAAA GGGGGATGTG
7451 CTGCAAGGCG ATTAAGTTGG GTAACGCCAG GTTTCGCCA GTCACGACGT
7501 TGTAAAACGA CGGCCAGTGA ATTACACCGG TGTGATCATG GGCCG

FIG 7B CONT.

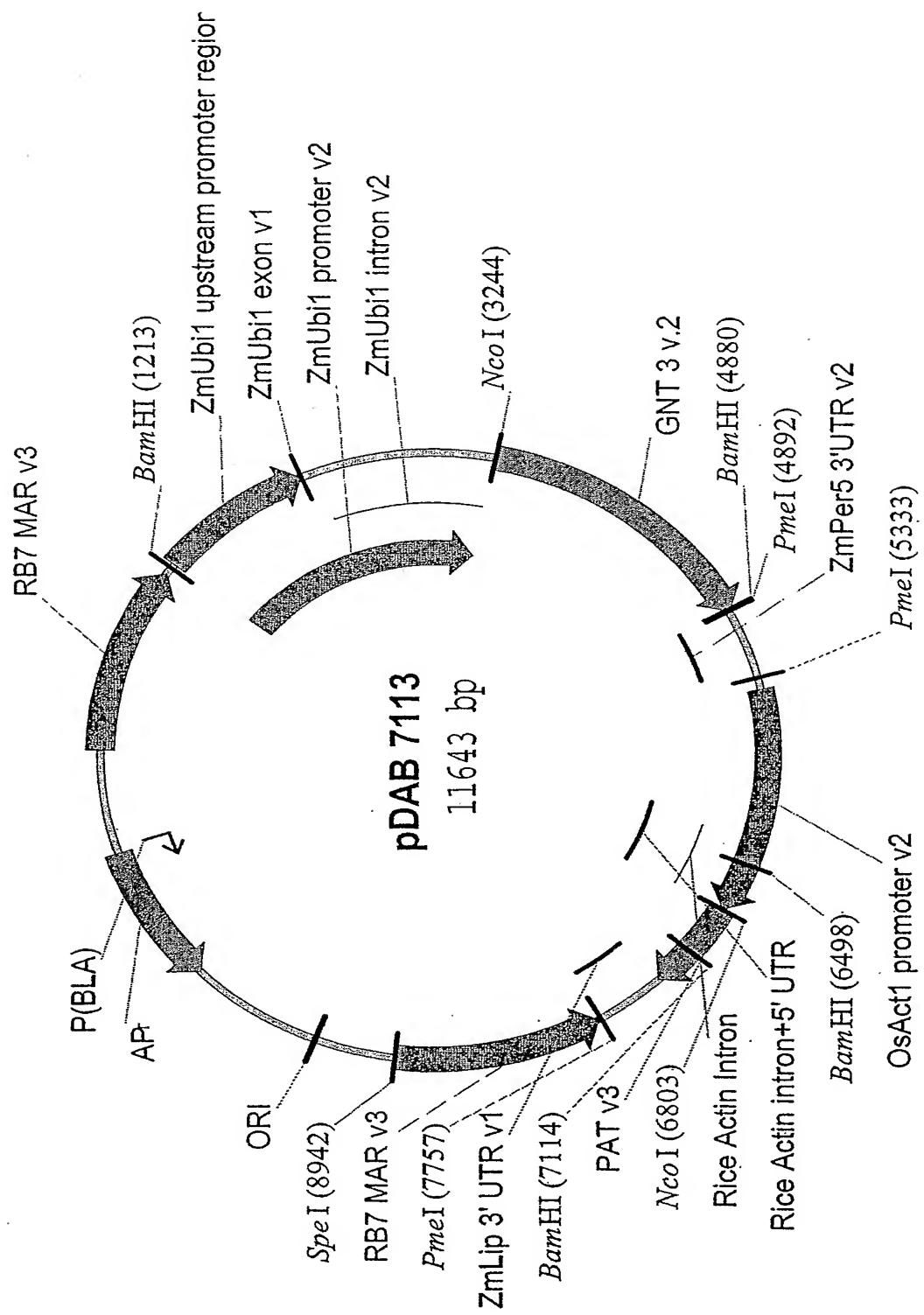


FIG. 8A

pDAB7113 11643bp

| Sequence | Feature |
|--|--|
| 1-1164bp | RB7 MAR v3 |
| 1165-1233
Linker | TGGCCACCGCTTAATTAAGGCGGCCATGCCCTGCAG
ATCCCCGGGATCCTCTAGAGTCGACCTGC |
| 1234-3224 | Maize Ubiquitin 1 promoter |
| 3224-4891 | GNT III v.2 |
| 3627 | C to replace G as reported in original sequence |
| 4892-4895
Linker | TAGTTT |
| 4896-5260 | Maize Per5 3'UTR v2 |
| 5261-5404
multiple
cloning sites | CGGCCGGCCTAGCTAGCCACGGTGGCCAGATCCACTAGG
GGCAAGCGGCCGCTTAATTAAATTAAATGTTAAACTA
GGAAATCCAAGCTGGGCTGCAGGTCAATCCCATTGCTT
TTGAAGCAGCTAACATTGATCTCTTT |
| 5405-6802 | Rice Actin 1 Promoter v2 |
| 6803-7358 | PAT v3 |
| 7359-7372
Linker | GGTACCCCTGAGCTC |
| 7373-7729 | Maize Lipase 3' UTR v1 |
| 7730-7770
Linker | GAATTCATATTCCTCCTGCAGGGTTAAACTTGCGTG
GC |
| 7771-8934 | RB7 MAR v3 |
| 8935-11643 | PUC19 |
| 9201-9225 | 3 FspI sites (TGCGCAA) with CG in between sites |
| 10164-11021 | Ampicillin resistance gene |
| 10454-10459 | TGCGCAA FspI |
| 11477-11482 | TGCGCAA FspI |

1 CGATTAaaaa CCCAATTATA TTTGGTCTAA TTTAGTTGG
 TATTGAGTAA AACAAATTG AACCAAACCA AAATATAAAT
 ATATAGTTTT TATATATATG
 101 CCTTTAAGAC TTTTATAGA ATTTCTTTA AAAAATATCT
 AGAAATATTT GCGACTCTTC TGGCATGTAA TATTCGTTA
 AATATGAAGT GCTCCATT
 201 TATTAACCTT AAATAATTGG TTGTACGATC ACTTTCTTAT
 CAAGTGTAC TAAAATGCGT CAATCTCTT GTTCTTCCAT
 ATTCAATATGT CAAAATCTAT
 301 CAAAATTCTT ATATATCTT TTCGAATTG AAGTGAATT
 TCGATAATT AAAATTAAAT AGAACATATC ATTATTTAGG
 TATCATATTG ATTTTATAC

401 TTAATTACTA AATTGGTTA ACTTTGAAAG TGTACATCAA
 CGAAAAATTA GTCAAACGAC TAAAATAAAT AAATATCATG
 TGTTATTAAG AAAATTCTCC
 501 TATAAGAATA TTTTAATAGA TCATATGTTT GTAAAAAAA
 TTAATTTTA CTAACACATA TATTTACTTA TCAAAAATTT
 GACAAAGTAA GATTAAAATA
 601 ATATTCATCT AACAAAAAAA AAACCAGAAA ATGCTGAAAA
 CCCGGCAAAA CCGAACCAAT CCAAACCGAT ATAGTTGGTT
 TGGTTTGATT TTGATATAAA
 701 CCGAACCAAC TCGGTCCATT TGCACCCCTA ATCATAATAG
 CTTTAATATT TCAAGATATT ATTAAGTTAA CGTTGTCAAT
 ATCCTGGAAA TTTTGCAAAA
 801 TGAATCAAGC CTATATGGCT GTAATATGAA TTTAAAAGCA
 GCTCGATGTG GTGGTAATAT GTAATTTACT TGATTCTAAA
 AAAATATCCC AAGTATTAAT
 901 AATTCTGCT AGGAAGAAGG TTAGCTACGA TTTACAGCAA
 AGCCAGAATA CAAAGAACCA TAAAGTGATT GAAGCTCGAA
 ATATACGAAG GAACAAATAT
 1001 TTTTAAAAAA ATACGCAATG ACTTGGAAACA AAAGAAAGTG
 ATATATTTT TGTTCTTAAA CAAGCATCCC CTCTAAAGAA
 TGGCAGTTT CCTTGCATG

PacI

~~~~~

## Ascl

~~~~~

1101 TAACTATTAT GCTCCCTTCG TTACAAAAT TTTGGACTAC
 TATTGGGAAT TCTTCTGAAA ATAGTGGCCA CCGCTTAATT
 AAGGCGCGCC ATGCCCTCTG

BamHI

~~~~~

1201 CAGATCCCCG GGGATCCTCT AGAGTCGACC TGCAGTGCAG  
 CGTGACCCGG TCGTCCCCT CTCTAGAGAT AATGAGCATT  
 GCATGTCTAA GTTATAAAA

1301 ATTACACAT ATTTTTTTG TCACACTTGT TTGAAGTGCA  
 GTTTATCTAT CTTTATACAT ATATTAAAC TTTAATCTAC  
 GAATAATATA ATCTATAGTA

1401 CTACAATAAT ATCAGTGTGTT TAGAGAATCA TATAAATGAA  
 CAGTTAGACA TGGTCTAAAG GACAATTGAG TATTTGACA  
 ACAGGACTCT ACAGTTTTAT

1501 CTTTTTAGTG TGCATGTGTT CTCCTTTTT TTTGCAAATA  
 GCTTCACCTA TATAATACTT CATCCATTAA ATTAGTACAT  
 CCATTTAGGG TTTAGGGTTA

1601 ATGGTTTTA TAGACTAATT TTTTAGTAC ATCTATTTA  
 TTCTATTTA GCCTCTAAAT TAAGAAAAT AAAACTCTAT  
 TTTAGTTTT TTATTTAATA  
 1701 ATTAGATAT AAAATAGAAT AAAATAAAGT GACTAAAAAT  
 TAAACAAATA CCCTTTAAGA AATTAAAAAA ACTAAGGAAA  
 CATTTTCTT GTTCGAGTA  
 1801 GATAATGCCA GCCTGTTAAA CGCCGTCGAC GAGTCTAACG  
 GACACCAACC AGCGAACCGAG CAGCGTCGCG TCGGGCCAAG  
 CGAACGAGAC GGCACGGCAT  
 1901 CTCTGTCGCT GCCTCTGGAC CCCTCTCGAG AGTTCCGCTC  
 CACC GTTGA CTTGCTCCGC TGTCGGCATC CAGAAATTGC  
 GTGGCGGAGC GGCAGACGTG  
 2001 AGCCGGCACG GCAGGGCGGCC TCCTCCTCCT CTCACGGCAC  
 GGCAGCTACG GGGGATTCCCT TTCCCACCGC TCCTTCGCTT  
 TCCCTTCCTC GCCCGCCGTA  
 2101 ATAAAATAGAC ACCCCCTCCA CACCCTCTT CCCCAACCTC  
 GTGTTGTTCG GAGCGCACAC ACACACAACC AGATCTCCCC  
 CAAATCCACC CGTCGGCACC  
 2201 TCCGCTTCAA GGTACGCCGC TCGTCCTCCC CCCCCCCCCC  
 TCTCTACCTT CTCTAGATCG GCGTCCGGT CCATGCATGG  
 TTAGGGCCCG GTAGTTCTAC  
 2301 TTCTGTTCAT GTTGTGTTA GATCCGTGTT TGTGTTAGAT  
 CCGTGCTGCT AGCGTTCGTA CACGGATGCG ACCTGTACGT  
 CAGACACGTT CTGATTGCTA  
 2401 ACTTGCCAGT GTTCTCTTT GGGGAATCCT GGGATGGCTC  
 TAGCCGTTCC GCAGACGGGA TCGATTTCAT GATTTTTTT  
 GTTTCGTTGC ATAGGGTTTG  
 2501 GTTTGCCCTT TTCTTTATT TCAATATATG CCGTGCACCT  
 GTTTGTCGGG TCATCTTTTC ATGCTTTTT TTGTCTGGT  
 TGTGATGATG TGGTCTGGTT  
 2601 GGGCGGTCGT TCTAGATCGG AGTAGAATTG TGTTCAAAC  
 TACCTGGTGG ATTATTAAT TTTGGATCTG TATGTGTGTG  
 CCATACATAT TCATAGTTAC  
 2701 GAATTGAAGA TGATGGATGG AAATATCGAT CTAGGATAGG  
 TATACATGTT GATGCGGGTT TTACTGATGC ATATACAGAG  
 ATGCTTTTG TTCGCTTGGT  
 2801 TGTGATGATG TGGTGTGGTT GGGCGGTCGT TCATTCGTT  
 TAGATCGGAG TAGAATACTG TTTCAAACTA CCTGGTGTAT  
 TTATTAATT TGGAACGTGA  
 2901 TGTGATGATG ATACATCTTC ATAGTTACGA GTTTAAGATG  
 GATGGAAATA TCGATCTAGG ATAGGTATAC ATGTTGATGT  
 GGGTTTACT GATGCATATA  
 3001 CATGATGGCA TATGCAGCAT CTATTCAAT GCTCTAACCT  
 TGAGTACCTA TCTATTATAA TAAACAAGTA TGTTTATAA  
 TTATTTGAT CTTGATATAC  
 3101 TTGGATGATG GCATATGCAG CAGCTATATG TGGATTTTT  
 TAGCCCTGCC TTCATACGCT ATTTATTGTC TTGGTACTGT  
 TTCTTTGTC GATGCTCACC

NcoI

FseI

~~~~~

~~~~~

3201 CTGTTACTTC TGAGGGTAC CCCCAGGGTC  
 GACCATGGTG ATGAGACGCT ACAAGCTCTT TCTCATGTTC  
 TGTATGGCCG GCCTGTGCCT

3301 CATCTCCTTC CTGCACTTCT TCAAGACCCT GTCTATGTC  
 ACCTTCCCCC GAGAACTGGC CTCCCTCAGC CCTAACCTGG  
 TGTCCAGCTT TTTCTGGAAC

3401 AATGCCCGG TCACGCCCA GGCCAGCCCC GAGCCAGGAG  
 GCCCTGACCT GCTGCGTACC CCACCTACT CCCACTCGCC  
 CCTGCTGCAG CCGCTGCCGC

SacI

~~~~~

3501 CCAGCAAGGC GGCGAGGAG CTCCACCGGG TGGACTTGGT
 GCTGCCGAG GACACCACCG AGTATTCGT GCGCACCAAG
 GCCGGCGGGCG TCTGCTTCAA

3601 ACCCGGCACC AAGATGCTGG AGAGGCCCGG CCCGGGACGG
 CGGGAGGAGA AGCCTGAGGG GGCCAACGGC TCCTCGGCC
 GGCGGCCACC CCGGTACCTC

3701 CTGAGCGCCC GGGAGCGCAC GGGGGGCCGA GGCGCCCGGC
 GCAAGTGGGT GGAGTGCCTG TGCCTGCCCG GCTGGCACGG
 ACCCAGCTGC GGCGTGCCCA

3801 CTGTGGTCA GTACTCCAAC CTGCCACCA AGGAGCGGCT
 GGTGCCAGG GAGGTGCCGC GCCGCGTCAT CAACGCCATC
 AACGTCAACC ACGAGTTCGA

NotI

~~~~~

3901 CCTGCTGGAC GTGCGCTTCC ACGAGCTGGG CGACGTGGTG  
 GACGCCCTTG TGGTGTGCGA GTCCAACCTTC ACGGCTTATG  
 GGGAGCCGCG GCCGCTCAAG

4001 TTCCGGGAGA TGCTGACCAA TGGCACCTTC GAGTACATCC  
 GCCACAAGGT GCTCTATGTC TTCCTGGACC ACTTCCCGCC  
 CGGCGGCCGG CAGGACGGCT

4101 GGATCGCCGA CGACTACCTG CGCACCTTCC TCACCCAGGA  
 CGGCGTCTCG CGGCTGCGCA ACCTGCGGCC CGACGACGTC  
 TTCATCATTG ACGATGCGGA

4201 CGAGATCCCG GCCCGTGACG GCGTCCTTT CCTCAAGCTC  
 TACGATGGCT GGACCGAGCC CTTCGCCTTC CACATGCGCA  
 AGTCGCTCTA CGGCTTCTTC

4301 TGGAAAGCAGC CGGGCACCCCT GGAGGTGGTG TCAGGCTGCA  
 CGGTGGACAT GCTGCAGGCA GTGTATGGGC TGGACGGCAT  
 CGGCCTGCGC CGCCGCCAGT

FIG 8B CONT.

4401 ACTACACCAT GCCCAACTTC AGACAGTATG AGAACCGCAC  
 CGGCCACATC CTGGTGCAGT GGTCGCTGGG CAGCCCCCTG  
 CACTTCGCCG GCTGGCACTG  
 4501 CTCCTGGTGC TTCACGCCCG AGGGCATCTA CTTCAAGCTC  
 GTGTCCGCCA AGAATGGCGA CTTCCCACGC TGGGGTGACT  
 ACGAGGACAA GCGGGACCTG  
 4601 AACTACATCC GCGGCCTGAT CCGCACCGGG GGCTGGTTCG  
 ACGGCACGCA GCAGGAGTAC CCGCCTGCAG ACCCCAGCGA  
 GCACATGTAT GCGCCAAGT  
 4701 ACCTGCTGAA GAACTACGAC CGGTTCCACT ACCTGCTGGA  
 CAACCCCTAC CAGGAGCCCA GGAGCACGGC GGCGGGCGGG  
 TGGCGCCACA GGGGTCCCGA

BamHI                  PmeI

~~~~~ ~~~~~

4801 GGGAAAGGCCG CCCGCCCGGG GCAAACCTGGA CGAGGCCGAA
 GTCGAACAAA AACTCATCTC AGAAGAGGAT CTGAATTAGG
 ATCCTAGGTT TAAACTGAGG
 4901 GCACTGAAGT CGCTTGATGT GCTGAATTGT TTGTGATGTT
 GGTGGCGTAT TTTGTTAAA TAAGTAAGCA TGGCTGTGAT
 TTTATCATAT GATCGATCTT
 5001 TGGGGTTTTA TTTAACACAT TGAAAATGT GTATCTATTAA
 ATAACCTCAAT GTATAAGATG TGTTCATTCT TCGGTTGCCA
 TAGATCTGCT TATTGACCT
 5101 GTGATGTTTT GACTCCAAAA ACCAAAATCA CAACTCAATA
 AACTCATGGA ATATGTCCAC CTGTTCTTG AAGAGTTCAT
 CTACCATTCC AGTTGGCATT

FseI

~~~~~

5201 TATCAGTGT GCAGCGGCAGC TGTGCTTTGT AACATAACAA  
 TTGTTCACGG CATATATCCA CGGCCGGCCT AGCTAGGCCAC  
 GGTGGCCAGA TCCACTAGGG

PacI

~~~~~

SwaI

~~~~~

NotI

PmeI

HindIII

~~~~~

~~~~~

5301 GCAAGCGGCC GCTTAATTAA ATTAAATGT TTAAACTAGG  
 AAATCCAAGC TTGGGCTGCA GGTCAATCCC ATTGCTTTG  
 AAGCAGCTCA ACATTGATCT

5401 CTTTCTCGAG GTCATTATA TGCTTGAGAA GAGAGTCGGG  
 ATAGTCCAAA ATAAAACAAA GGTAAGATTA CCTGGTCAAA  
 AGTAAAACA TCAGTTAAAA  
 5501 GGTGGTATAA AGTAAAATAT CGGTAATAAA AGGTGGCCCA  
 AAGTGAATT TACTCTTTC TACTATTATA AAAATTGAGG  
 ATGTTTTGT CGGTACTTTG  
 5601 ATACGTCATT TTTGTATGAA TTGGTTTTA AGTTTATTG  
 CTTTGGAAA TGCATATCTG TATTGAGTC GGGTTTAAG  
 TTCGTTTGCT TTTGTAAATA  
 5701 CAGAGGGATT TGTATAAGAA ATATCTTTAA AAAAACCCAT  
 ATGCTAATTG GACATAATTG TTGAGAAAAA TATATATTCA  
 GGCAGATTCT CACAATGAAC  
 5801 AATAATAAGA TTAAAATAGC TTTCCCCCGT TGCAGCGCAT  
 GGGTATTGTT TCTAGTAAAA ATAAAAGATA AACTTAGACT  
 CAAAACATTG ACAAAACAA  
 5901 CCCCTAAAGT TCCTAAAGCC CAAAGTGCTA TCCACGATCC  
 ATAGCAAGCC CAGCCCAACC CAACCCAACC CAACCCACCC  
 CAGTCCAGCC AACTGGACAA  
 6001 TAGTCTCCAC ACCCCCCCAC TATCACCGTG AGTTGTCCGC  
 ACGCACCGCA CGTCTCGCAG CCAAAAAAAA AAAAGAAAG  
 AAAAAAAAGA AAAAGAAAAA

## FseI

~~~~~

6101 ACAGCAGGTG GGTCCGGGTC GTGGGGGCCG GAAACGCGAG
 GAGGATCGCG AGCCAGCGAC GAGGCCGGCC CTCCCTCCGC
 TTCCAAAGAA ACGCCCCCA
 6201 TCGCCACTAT ATACATAACCC CCCCCTCTCC TCCCATCCCC
 CCAACCTAC CACCACCACC ACCACCACCT CCACCTCCTC
 CCCCCTCGCT GCCGGACGAC
 6301 GCCTCCCCCC TCCCCCTCCG CCGCCGCCGC GCCGGTAACC
 ACCCCGCCCG TCTCCTCTTT CTTTCTCCGT TTTTTTTTC
 CGTCTCGGTC TCGATCTTTG

BamHI

~~~~

6401 GCCTTGGTAG TTTGGGTGGG CGAGAGGGCGG CTTCGTGC  
 GCCCAGATCG GTGCGCGGGA GGGGGGGGAT CTCGCGGCTG  
 GGGCTCTCGC CGGCGTGGAT

## BamHI

~~

6501 CCGGCCCGGA TCTCGCGGGG AATGGGGCTC TCGGATGTAG  
 ATCTGCGATC CGCCGTTGTT GGGGGAGATG ATGGGGGGTT  
 TAAAATTCC GCCATGCTAA

6601 ACAAGATCAG GAAGAGGGGA AAAGGGCACT ATGGTTTATA  
       TTTTTATATA TTTCTGCTGC TTCGTCAGGC TTAGATGTGC  
       TAGATCTTTC TTTCTTCTTT  
 6701 TTGTGGGTAG AATTGAATC CCTCAGCATT GTTCATCGGT  
       AGTTTTCTT TTCATGATT GTGACAAATG CAGCCTCGTG  
       CGGAGCTTT TTGTAGGTAG

NcoI  
 ~~~~~

6801 ACCATGGCTT CTCCGGAGAG GAGACCAGTT GAGATTAGGC
 CAGCTACAGC AGCTGATATG GCCGCGGTTT GTGATATCGT
 TAACCATTAC ATTGAGACGT
 6901 CTACAGTGAA CTTTAGGACA GAGCCACAAA CACCACAAGA
 GTGGATTGAT GATCTAGAGA GGTTGCAAGA TAGATACCCT
 TGGTTGGTTG CTGAGGTTGA
 7001 GGGTGTGTTG GCTGGTATTG CTTACGCTGG GCCCTGGAAG
 GCTAGGAACCG CTTACGATTG GACAGTTGAG AGTACTGTTT
 ACGTGTCACA TAGGCATCAA

BamHI
 ~~~~~

7101 AGGTTGGGCC TAGGATCCAC ATTGTACACA CATTGCTTA  
       AGTCTATGGA GGCGCAAGGT TTTAAGTCTG TGGTTGCTGT  
       TATAGGCCTT CCAAACGATC  
 7201 CATCTGTTAG GTTGCATGAG GCTTTGGGAT ACACAGCCCG  
       GGGTACATTG CGCGCAGCTG GATACAAGCA TGGTGGATGG  
       CATGATGTTG GTTTTGGCA

SacI

~~~~~

7301 AAGGGATTTT GAGTTGCCAG CTCCTCCAAG GCCAGTTAGG
 CCAGTTACCC AGATCTGAGG TACCCCTGAGC TCGGTCGCAG
 CGTGTGCGTG TCCGTCGTAC

FseI

~~~~~

7401 GTTCTGGCCG GCCGGGCCTT GGGCGCGCGA TCAGAAGCGT  
       TGCCTGGCG TGTGTGTGCT TCTGGTTGC TTTAATTAA  
       CCAAGTTGT TTCAAGGTGG

7501 ATCGCGTGGT CAAGGCCGT GTGCTTTAAA GACCCACCGG  
       CACTGGCAGT GAGTGTGCT GCTTGTGTAG GCTTTGGTAC  
       GTATGGGCTT TATTGCTTC

7601 TGGATGTTGT GTACTACTTG GGTTTGGTGA ATTATTATGA  
       GCAGTTGCGT ATTGTAATTC AGCTGGGCTA CCTGGACATT  
       GTTATGTATT AATAAATGCT

PmeI

~~~~~

FIG 8B CONT.

7701 TTGCTTTCTT CAAAGATCT TTAAGTGCTG AATTCAATT
 TCCTCCTGCA GGGTTAACAC TTGCCGTGGC CTATTTCA
 AAGAATTCCC AATAGTAGTC
 7801 CAAAATTTT GTAACGAAGG GAGCATAATA GTTACATGCA
 AAGGAAAACT GCCATTCTT AGAGGGGATG CTTGTTAAG
 AACAAAAAAT ATATCACTTT
 7901 CTTTGTTCC AAGTCATTGC GTATTTTTT AAAAATATT
 GTTCCTTCGT ATATTCGAG CTTCAATCAC TTTATGGTC
 TTTGTATTCT GGCTTGCTG
 8001 TAAATCGTAG CTAACCTCT TCCTAGCAGA AATTATTAAT
 ACTTGGGATA TTTTTTTAGA ATCAAGTAAA TTACATATTA
 CCACCACATC GAGCTGCTTT
 8101 TAAATTCAATA TTACAGCCAT ATAGGCTTGA TTCATTTC
 AAAATTCCA GGATATTGAC AACGTTAACT TAATAATATC
 TTGAAATATT AAAGCTATTA
 8201 TGATTAGGGG TGCAAATGGA CCGAGTTGGT TCGGTTATA
 TCAAAATCAA ACCAAACCAA CTATATCGGT TTGGATTGGT
 TCGGTTTGC CGGGTTTCA
 8301 GCATTTCTG GTTTTTTTT TGTTAGATGA ATATTATTT
 AATCTTACTT TGTCAAATT TTGATAAGTA AATATATGTG
 TTAGTAAAAA TTAATTTTT
 8401 TTACAAACAT ATGATCTATT AAAATATTCT TATAGGAGAA
 TTTCTTAAT AACACATGAT ATTTATTTAT TTTAGTCGTT
 TGACTAATT TTCGTTGATG
 8501 TACACTTTCA AAGTTAACCA AATTAGTAA TTAAGTATAA
 AAATCAATAT GATACTAAA TAATGATATG TTCTATTTAA
 TTTAAATTAA TCGAAATTTC
 8601 ACTTCAAATT CGAAAAAGAT ATATAAGAAT TTTGATAGAT
 TTTGACATAT GAATATGGAA GAACAAAGAG ATTGACGCAT
 TTTAGTAACA CTTGATAAGA
 8701 AAGTGATCGT ACAACCAATT ATTTAAAGTT AATAAAAATG
 GAGCACTTCA TATTTAACGA AATATTACAT GCCAGAAGAG
 TCGCAAATAT TTCTAGATAT
 8801 TTTTAAAGA AAATTCTATA AAAAGTCTTA AAGGCATATA
 TATAAAAACAT ATATATTTAT ATTTGGTTT GGTCGAATT
 TGTTTACTC AATACCAAAC
 8901 TAAATTAGAC CAAATATAAT TGGGTTTTA ATCGCGGCC
 ACTAGTCACC GGTGTAGCTT GGCCTAATCA TGGTCATAGC
 TGTTTCTGT GTGAAATTGT
 9001 TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT
 CACATTAATT GCGTTGCGCT
 9101 CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
 GCATTAATGA ATCGGCCAAC GCGCGGGGAG AGGCGGTTG
 CGTATTGGGC GCTCTTCCGC
 9201 TGGCACGCT GCGCACGCTG CGCACGCTTC CTCGCTCACT
 GACTCGCTGC GCTCGGTGCGT TCGGCTGCGG CGAGCGGTAT
 CAGCTCACTC AAAGGCGGTA

9301 ATACGGTTAT CCACAGAAC AGGGATAAC GCAGGAAAGA
 ACATGTGAGC AAAAGGCCAG CAAAAGGCCA GGAACCGTAA
 AAAGGCCGCG TTGCTGGCGT
 9401 TTTCCATAG GCTCCGCCCG CCTGACGAGC ATCACAAAAA
 TCGACGCTCA AGTCAGAGGT GGCGAAACCC GACAGGACTA
 TAAAGATACC AGGCGTTTCC
 9501 CCCTGGAAGC TCCCTCGTGC GCTCTCCTGT TCCGACCCCTG
 CCGCTTACCG GATAACCTGTC CGCCTTCTC CCTTCGGGAA
 GCGTGGCGCT TTCTCATAGC
 9601 TCACGCTGTA GGTATCTCAG TTCGGTGTAG GTCGTTCGCT
 CCAAGCTGGG CTGTGTGCAC GAACCCCCCG TTCAGCCGA
 CCGCTGCGCC TTATCCGGTA
 9701 ACTATCGTCT TGAGTCCAAC CCGGTAAGAC ACGACTTATC
 GCCACTGGCA GCAGCCACTG GTAACAGGAT TAGCAGAGCG
 AGGTATGTAG GCGGTGCTAC
 9801 AGAGTTCTTG AAGTGGTGGC CTAACTACGG CTACACTAGA
 AGGACAGTAT TTGGTATCTG CGCTCTGCTG AAGCCAGTTA
 CCTTCGGAAA AAGAGTTGGT
 9901 AGCTCTTGAT CCGGCAAACA AACCAACCGCT GGTAGCGGTG
 GTTTTTTGT TTGCAAGCAG CAGATTACGC GCAGAAAAAA
 AGGATCTCAA GAAGATCCTT
 10001 TGATCTTTTC TACGGGTCT GACGCTCAGT GGAACGAAAAA
 CTCACGTTAA GGGATTTGG TCATGAGATT ATCAAAAGG
 ATCTTCACCT AGATCCTTTT
 10101 AAATTAAAAAA TGAAGTTTTA AATCAATCTA AAGTATATAT
 GAGTAAACTT GGTCTGACAG TTACCAATGC TTAATCAGTG
 AGGCACCTAT CTCAGCGATC
 10201 TGTCTATTTTC GTTCATCCAT AGTTGCCTGA CTCCCCGTG
 TGTAGATAAC TACGATACGG GAGGGCTTAC CATCTGGCCC
 CAGTGCTGCA ATGATACCGC
 10301 GAGACCCACG CTCACCGGCT CCAGATTTAT CAGCAATAAA
 CCAGCCAGCC GGAAGGGCCG AGCGCAGAAG TGGTCCTGCA
 ACTTTATCCG CCTCCATCCA
 10401 GTCTATTAAT TGTTGCCGGG AAGCTAGAGT AAGTAGTTCG
 CCAGTTAATA GTTTGCGCAA CGTTGTTGCC ATTGCTACAG
 GCATCGTGGT GTCACGCTCG
 10501 TCGTTGGTA TGGCTTCATT CAGCTCCGGT TCCCAACGAT
 CAAGGCGAGT TACATGATCC CCCATGTTGT GCAAAAGC
 GGTTAGCTCC TTCGGTCCTC
 10601 CGATCGTTGT CAGAAGTAAG TTGGCCGCAG TGTTATCACT
 CATGGTTATG GCAGCACTGC ATAATTCTCT TACTGTCATG
 CCATCCGTAA GATGCTTTTC
 10701 TGTGACTGGT GAGTACTCAA CCAAGTCATT CTGAGAATAG
 TGTATGCGGC GACCGAGTTG CTCTTGCCCG GCGTCAATAC
 GGGATAATAC CGCGCCACAT
 10801 AGCAGAACTT TAAAAGTGCT CATCATTGGA, AACGTTCTT
 CGGGCGAAA ACTCTCAAGG ATCTTACCGC TGTTGAGATC
 CAGTTCGATG TAACCCACTC

10901 GTGCACCCAA CTGATCTTCA GCATCTTTA CTTTCACCAAG
CGTTCTGGG TGAGCAAAAA CAGGAAGGCA AAATGCCGCA
AAAAAGGGAA TAAGGGCGAC
11001 ACGGAAATGT TGAATACTCA TACTCTTCCT TTTTCAATAT
TATTGAAGCA TTTATCAGGG TTATTGTCTC ATGAGCGGAT
ACATATTGAG ATGTATTTAG
11101 AAAAATAAAC AAATAGGGGT TCCGCGCACA TTTCCCCGAA
AAGTGCCACC TGACGTCTAA GAAACCATTA TTATCATGAC
ATTAACCTAT AAAAATAGGC
11201 GTATCACGAG GCCCTTCGT CTCGCGCGTT TCGGTGATGA
CGGTGAAAAC CTCTGACACA TGCAGCTCCC GGAGACGGTC
ACAGCTTGTC TGTAAGCGGA
11301 TGCCGGGAGC AGACAAGCCC GTCAGGGCGC GTCAGCGGGT
GTTGGCGGGT GTCGGGGCTG GCTTAACTAT GCGGCATCAG
AGCAGATTGT ACTGAGAGTG
11401 CACCATATGC GGTGTGAAAT ACCGCACAGA TGCGTAAGGA
GAAAATACCG CATCAGGCAGC CATTGCCAT TCAGGCTGCG
CAACTGTTGG GAAGGGCGAT
11501 CGGTGCGGGC CTCTTCGCTA TTACGCCAGC TGGCGAAAGG
GGGATGTGCT GCAAGGCGAT TAAGTTGGGT AACGCCAGGG
TTTCCCAGT CACGACGTTG
11601 TAAAACGACG GCCAGTGAAT TACACCGGTG TGATCATGGG CCG

FIG 8B CONT.

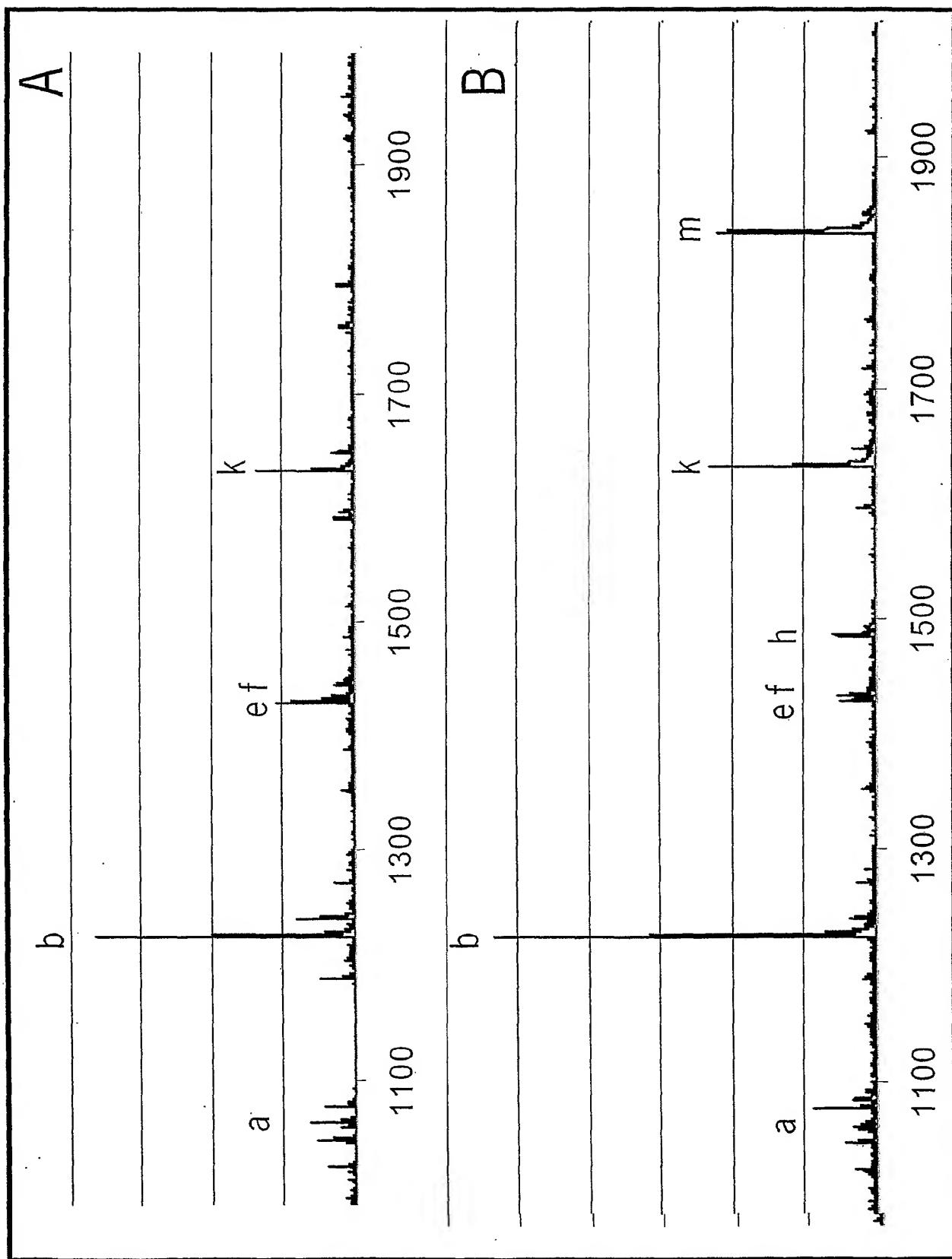


FIG. 9

atgaagatgagacgctacaagctcttcatgttcttatggccggcgtgcctcatctccttcgtcacttcaagaccct
gtcctatgtcaccccccggagaactggcccccctcagccctaaccctgggtccagctttcttggaaacaatgccccggta
cgccccaggccagccccggccaggaggccctgacctgctcgtaaaaaactactccactgccccgtcagcc
gctgcccccagcaaggccggcggaggagctccaccgggtggacttgggtctgcccggaggacaccacccggatatttgt
gcccaccaaggccggcggcgtcgttcaaaccggcaccatggatgtggagaggccggggacggccggag
gagaaggctgaggggggccaacggctcccgccccggccacccggfacccttgcgacggccggagcgcacgg
ggggccgaggccggcgcgaagtgggtggagtgcgtgtgcctgcccggctggcacggacccagctgcggcgttg
cactgtgtgcagacttcaaccctgcccaccaaggagccggctgggtggccaggatgcgcgcgtcatcaacgc
catcaacgtcaaccacgagttcgacctgacgtgcgttccacgagctggcgacgtggtgacgccttggat
cgagtccaaacttacggcttatggggagccggccgcataagttccggagatgtgaccaatggcacccatcgact
tccggccacaagggtgttatgttccctggaccacttcccgccccggccggcaggacggctggatgcgcgcact
ctgcgcacccctcaccaggacggcgtctcgccgtgcacccctggccggccacgcgtctcatattgacatgc
ggacgagatcccgcccgfagccgtcccttcccaagcttacgatggctggaccgagccctgcctccacatgc
caagtgcgttacggcttcttgcgaagcagccggcaccctggagggtgggtcaggctgcacggtgacatgc
caggatgggtggacggcattccgcgttgcacttgcgttgcacttgcgttgcacttgcgttgcact
accggccacatccgttgcacttgcgttgcacttgcgttgcacttgcgttgcact
cgagggcatctacttcaagctcggtccggccagaatggcgacttccacgctggggactacgaggacaagg
ctgaactacatccgcggcctgatccgcaccggggctggatccgcacggcactgcaggataccgc
agcgagcacatgtatgcgcccacttgcacttgcacttgcgttgcact
gcccaggagcacggccggccgggtggcgccacagggtcccaggatggaaaggccggccggggcaaactgg
acgaggcggaagtcttag

FIG. 10

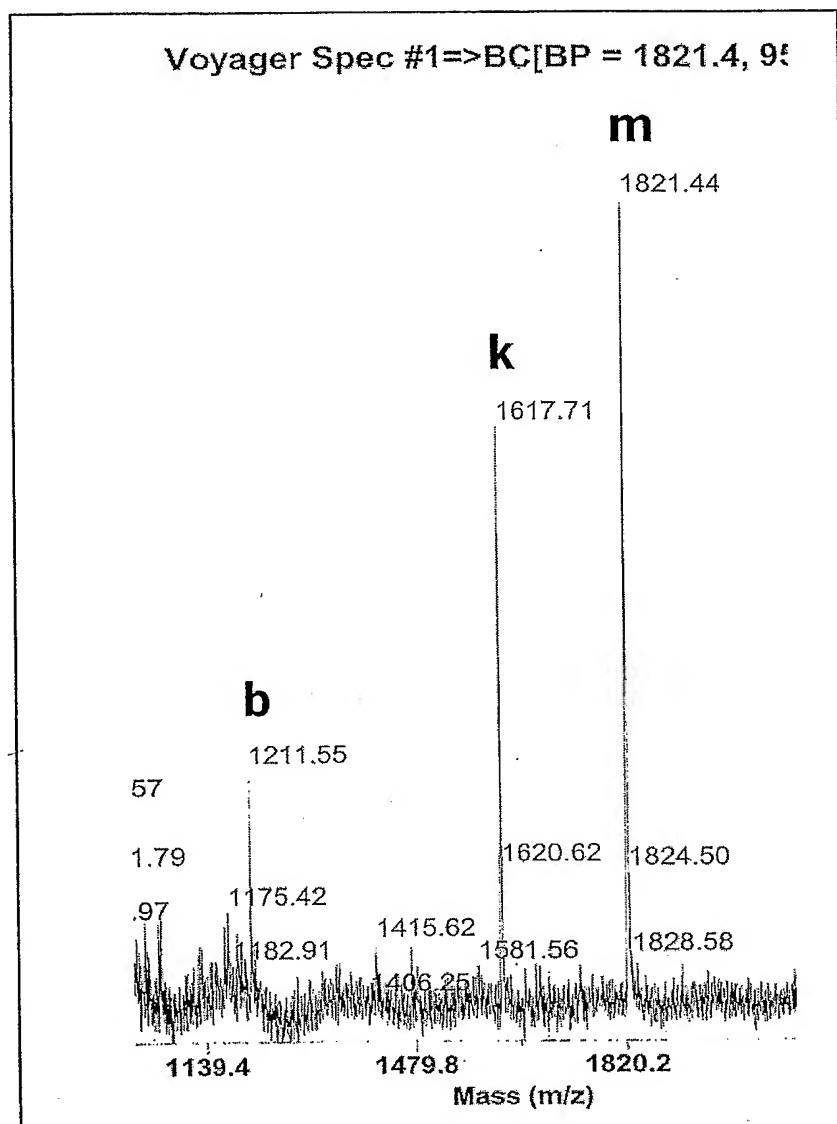


FIG 11

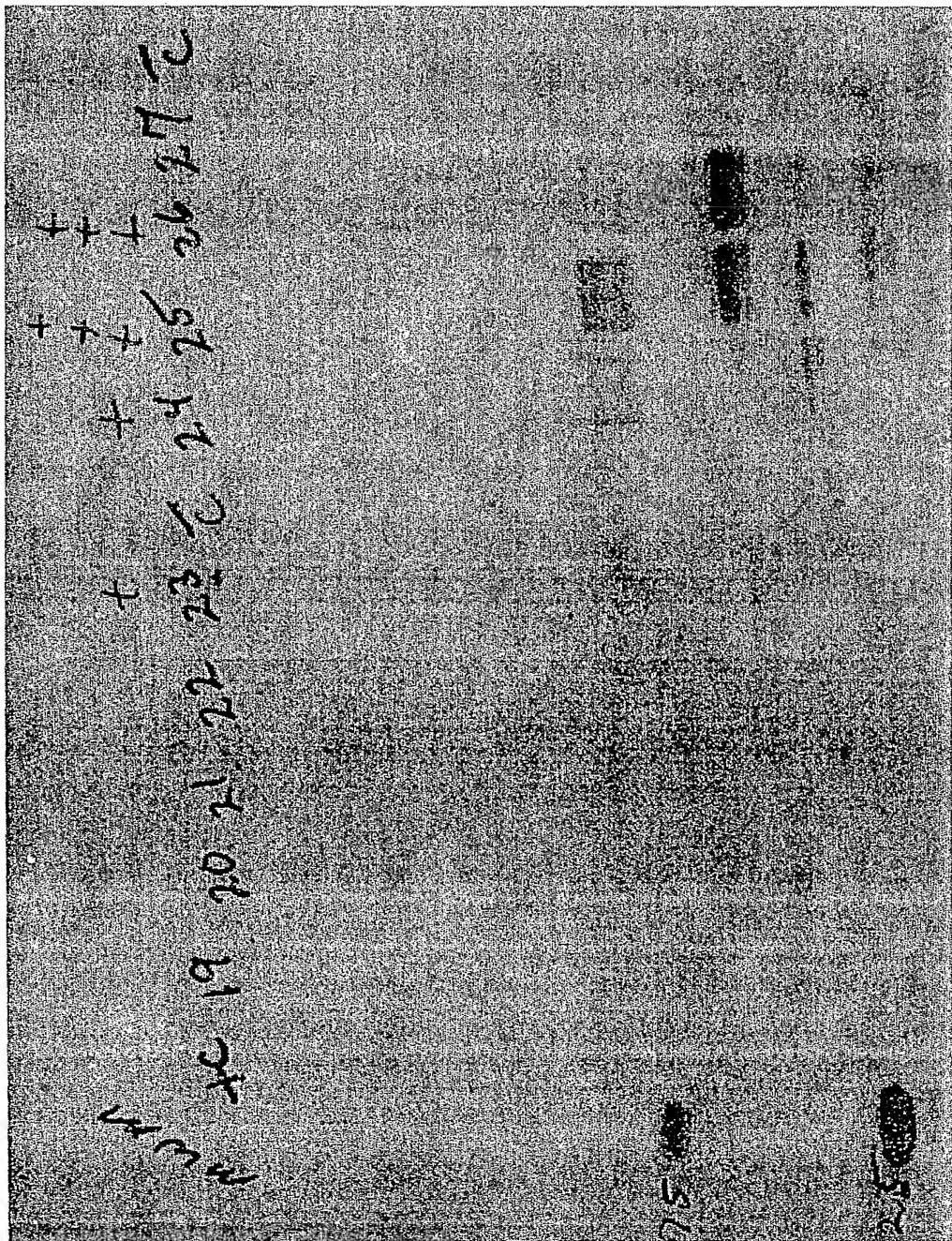


FIG 12

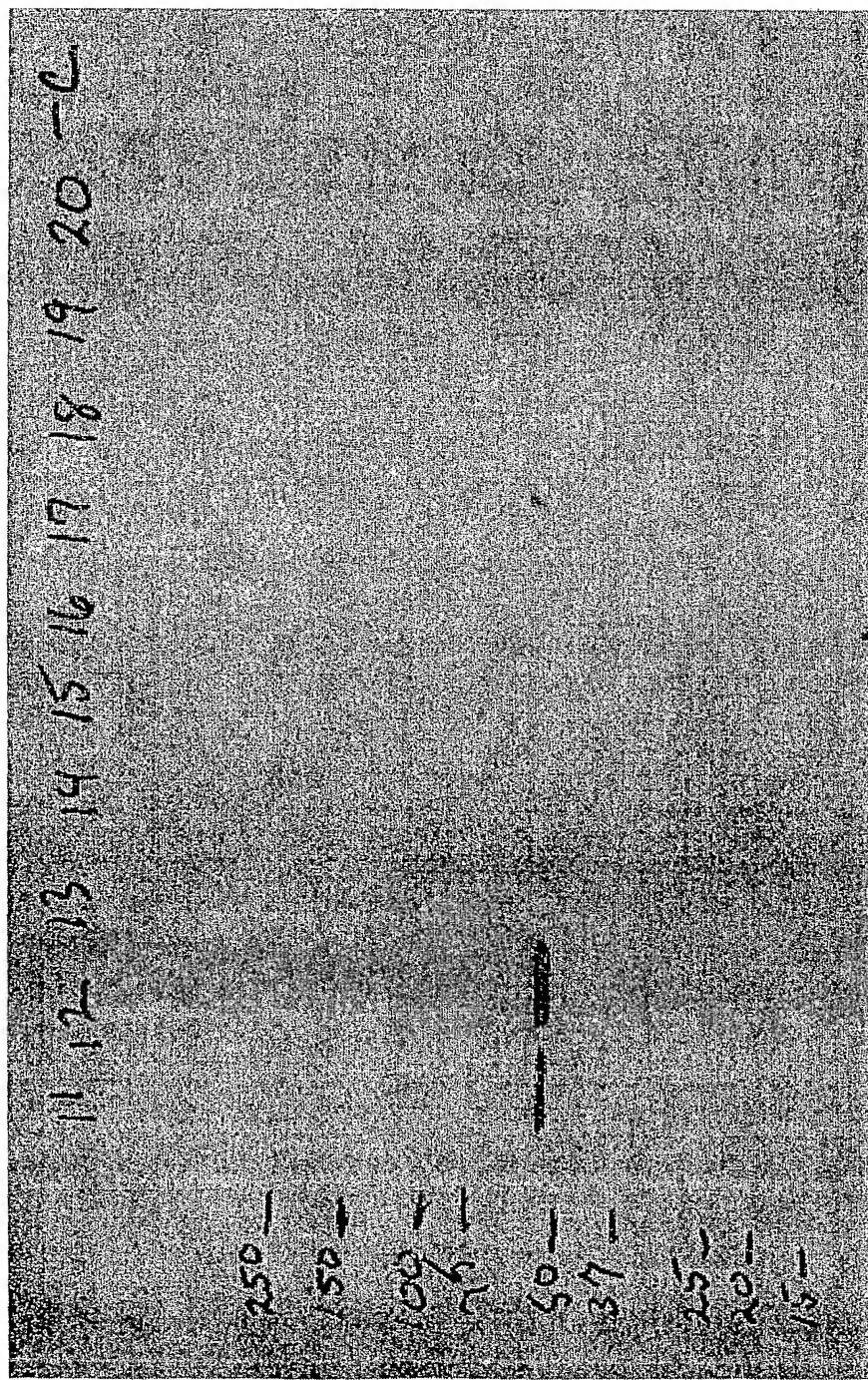


FIG 13

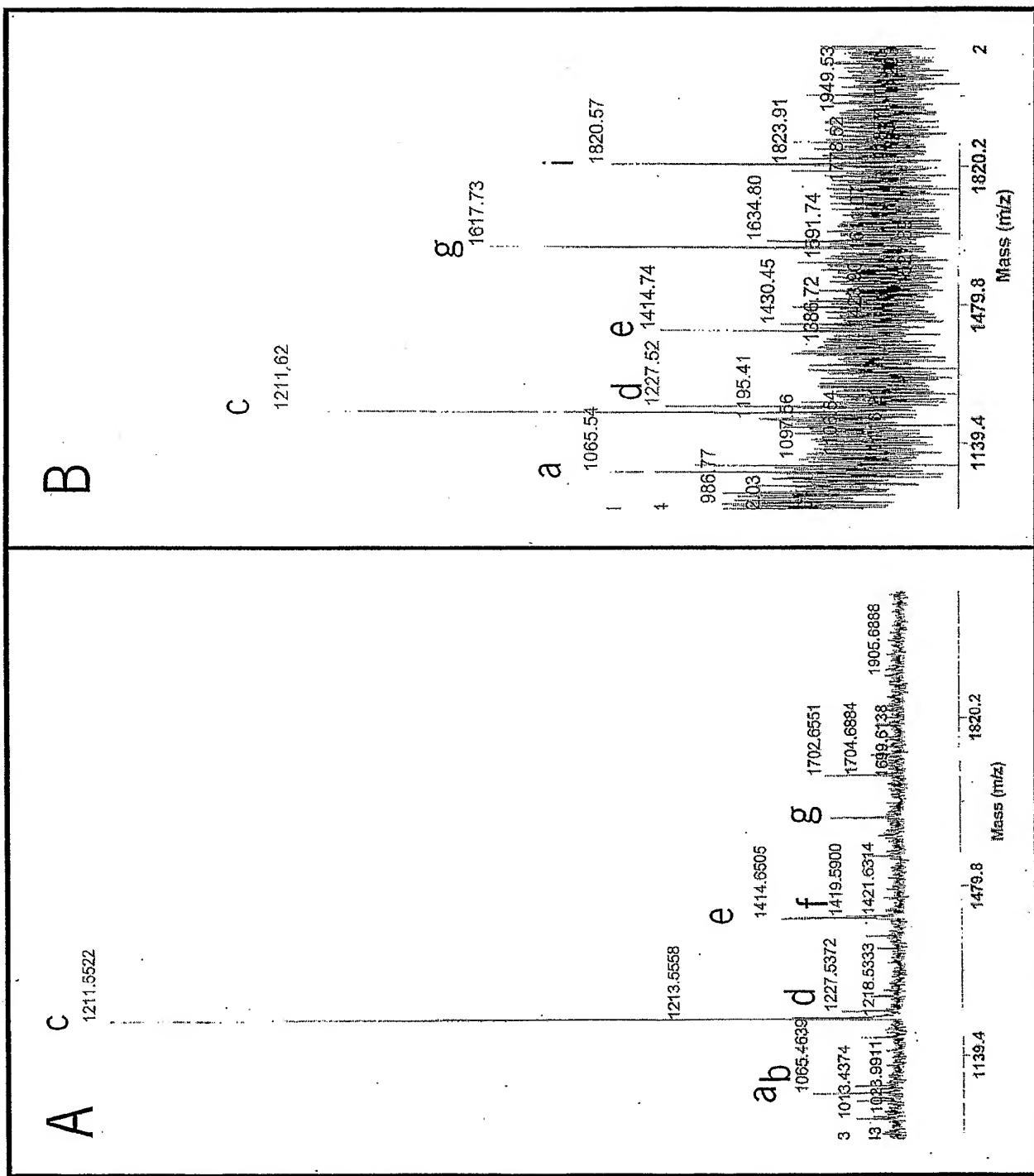


FIG. 14

SEQUENCE LISTING

<110> PLANT RESEARCH INTERNATIONAL BV
BAKKER, Hendrikus A.C.
FLORACK, Dionisius E.A.
BOSCH, Hendrik J.

<120> GNTIII expression in plants

<130> 62862A - P033313W0

<150> US-60/365,769

<151> 2002-03-19

<150> US-60/368,047

<151> 2002-03-26

<160> 27

<170> PatentIn version 3.2

<210> 1

<211> 1642

<232> DNA

<213> Homo sapiens

<400> 1
ccatgggtgat gagacgcgtac aagctcttc tcatgttctg tatggccggc ctgtgcctca
tctccttcctt gcacttcttc aagaccctgt cttatgtcac cttcccccgaa gaactggccc
ccctcagccc taacctggtg tccagcttt tcttggaaaca tgccccggtc acgccccagg
ccagccccga gccaggaggc cctgacccgtc tgctgatcccc actctactcc cactcgcccc
tgctgcagcc gctgccgccc agcaaggcg ccgaggagct ccaccgggtg gacttgggtgc
tgcccgagga caccaccgag tatttcgtgc gcaccaaggc cggcggcgtc tgcttcaa
ccggcacccaa gatgctggag aggccgcccc cgggacggcc ggaggagaag cctgagggggg
ccaacggttc ctcggcccccgg cgcccccccccc ggtacccctt gagcggccgg gagcgcacgg
ggggccgagg cggccggcgc aagtgggtgg atgtcggtg cctggccggc tggcagcggac
ccagctgggg cgtggccact gtggtgcagt actccaaacctt gcccaccaag gaggcgctgg
tgcccgagga ggtgccgccc cgcgtcatca acggccatcaa cgtcaaccac gagttcgacc
tgctggacgt gcgtttccac gagctggggc acgtgggtgg cgccttggc gtgtgcgaggt
ccaacttcac ggcttatggg gagccggcgc cgctcaagtt ccgggagatg ctgaccaatg
gcacccctgaa gtacatccgc cacaagggtgc tctatgtctt cctggaccac ttcccggcc
gcggccggca ggacggctgg atcgccgacg actacccgtcg caccttcctc accccaggacg
gcgtctcgcg gctgcgcaac ctgcggcccc acgacgtctt catattgac gatgcggacg
agatccggc cctgtacggc gtcctttcc tcaagctcta cgatggctgg accggccct
tcgccttcca catgcgcgaag tcgtcttaacg gttcttctg gaaagcggccg ggcacccctgg
agggtgttc aggtgcacacg gtggacatgc tgcaggcagt gtatggctg gacggcatcc
gcctgcggcg ccgcccagtac tacaccatgc ccaacttcag acagtatgag aaccgcaccc
gccacatcct ggtgcagtgg tcgctggca gccccctgca cttcggccggc tggcactgt
cctgggtctt cacgccccgag ggcataact tcaagctcggt gtccggcccg aatggcgact
tcccacgctg ggtgactac gaggacaagg gggacctgaa ctacatccgc ggcctgtatcc
gcaccggggg ctggttcgac ggcacgcgc aggatgtaccc gcctgcagac cccagcggac
acatgtatgc gcccaagtac tcgtcttaaaa actacgaccc gttccactac ctgctggaca
acccttacca ggagccccagg agcacggcg cgggggggtt ggcacccagg ggtcccgagg
gaaggccggcc cggccgggggg aaactggacg aggccggaaagt cgaacaaaaaa ctcatctcag
aaqaggatct gaattaggat cc

<210> 2

<211> 544

<212> PRT

<213> Homo sapiens

<400> 2

Met Val Met Arg Arg Tyr Lys Leu Phe Leu Met Phe Cys Met Ala Gly
1 5 10 15

Leu Cys Leu Ile Ser Phe Leu His Phe Phe Lys Thr Leu Ser Tyr Val
20 25 30

2/15

Thr Phe Pro Arg Glu Leu Ala Ser Leu Ser Pro Asn Leu Val Ser Ser
 35 40 45

Phe Phe Trp Asn Asn Ala Pro Val Thr Pro Gln Ala Ser Pro Glu Pro
 50 55 60

Gly Gly Pro Asp Leu Leu Arg Thr Pro Leu Tyr Ser His Ser Pro Leu
 65 70 75 80

Leu Gln Pro Leu Pro Pro Ser Lys Ala Ala Glu Glu Leu His Arg Val
 85 90 95

Asp Leu Val Leu Pro Glu Asp Thr Thr Glu Tyr Phe Val Arg Thr Lys
 100 105 110

Ala Gly Gly Val Cys Phe Lys Pro Gly Thr Lys Met Leu Glu Arg Pro
 115 120 125

Pro Pro Gly Arg Pro Glu Glu Lys Pro Glu Gly Ala Asn Gly Ser Ser
 130 135 140

Ala Arg Arg Pro Pro Arg Tyr Leu Leu Ser Ala Arg Glu Arg Thr Gly
 145 150 155 160

Gly Arg Gly Ala Arg Arg Lys Trp Val Glu Cys Val Cys Leu Pro Gly
 165 170 175

Trp His Gly Pro Ser Cys Gly Val Pro Thr Val Val Gln Tyr Ser Asn
 180 185 190

Leu Pro Thr Lys Glu Arg Leu Val Pro Arg Glu Val Pro Arg Arg Val
 195 200 205

Ile Asn Ala Ile Asn Val Asn His Glu Phe Asp Leu Leu Asp Val Arg
 210 215 220

Phe His Glu Leu Gly Asp Val Val Asp Ala Phe Val Val Cys Glu Ser
 225 230 235 240

Asn Phe Thr Ala Tyr Gly Glu Pro Arg Pro Leu Lys Phe Arg Glu Met
 245 250 255

Leu Thr Asn Gly Thr Phe Glu Tyr Ile Arg His Lys Val Leu Tyr Val
 260 265 270

Phe Leu Asp His Phe Pro Pro Gly Gly Arg Gln Asp Gly Trp Ile Ala
 275 280 285

Asp Asp Tyr Leu Arg Thr Phe Leu Thr Gln Asp Gly Val Ser Arg Leu
 290 295 300

Arg Asn Leu Arg Pro Asp Asp Val Phe Ile Ile Asp Asp Ala Asp Glu
 305 310 315 320

Ile Pro Ala Arg Asp Gly Val Leu Phe Leu Lys Leu Tyr Asp Gly Trp
 325 330 335

Thr Glu Pro Phe Ala Phe His Met Arg Lys Ser Leu Tyr Gly Phe Phe
 340 345 350

Trp Lys Gln Pro Gly Thr Leu Glu Val Val Ser Gly Cys Thr Val Asp
 355 360 365

Met Leu Gln Ala Val Tyr Gly Leu Asp Gly Ile Arg Leu Arg Arg Arg
 370 375 380

Gln Tyr Tyr Thr Met Pro Asn Phe Arg Gln Tyr Glu Asn Arg Thr Gly
 385 390 395 400

His Ile Leu Val Gln Trp Ser Leu Gly Ser Pro Leu His Phe Ala Gly
 405 410 415

3/15

Trp	His	Cys	Ser	Trp	Cys	Phe	Thr	Pro	Glu	Gly	Ile	Tyr	Phe	Lys	Leu
				420					425				430		
Val	Ser	Ala	Gln	Asn	Gly	Asp	Phe	Pro	Arg	Trp	Gly	Asp	Tyr	Glu	Asp
				435			440				445				
Lys	Arg	Asp	Leu	Asn	Tyr	Ile	Arg	Gly	Leu	Ile	Arg	Thr	Gly	Gly	Trp
					450		455				460				
Phe	Asp	Gly	Thr	Gln	Gln	Glu	Tyr	Pro	Pro	Ala	Asp	Pro	Ser	Glu	His
					465	470				475				480	
Met	Tyr	Ala	Pro	Lys	Tyr	Leu	Leu	Lys	Asn	Tyr	Asp	Arg	Phe	His	Tyr
					485				490				495		
Leu	Leu	Asp	Asn	Pro	Tyr	Gln	Glu	Pro	Arg	Ser	Thr	Ala	Ala	Gly	Gly
				500			505					510			
Trp	Arg	His	Arg	Gly	Pro	Glu	Gly	Arg	Pro	Pro	Ala	Arg	Gly	Lys	Leu
					515		520				525				
Asp	Glu	Ala	Glu	Val	Glu	Gln	Lys	Leu	Ile	Ser	Glu	Glu	Asp	Leu	Asn
					530		535				540				

```
<210> E
<211> E1
<212> DNA
<213> Artificial Sequence
```

<220>
<223> Synthetic

<400> 3
atactcgagt taacaatgaa gatgagacgc t

三

```
<210> 4
<211> 35
<212> DNA
<213> Artificial Sequence
```

<220>
<223> Synthetic

<400> 4
tatggatcct aattcagatc ctcttctgag atgag

35

```
<210> 5
<211> 20
<212> DNA
<213> Artificial Sequence
```

<220>
<223> Synthetic

<400> 5
ccatggtgat gagacgctac

20

```
<210> 6
<211> 32
<212> DNA
<213> Artificial Sequence
```

<220>
<223> Synthetic

<400> 6

gtttaaacct aggatcctaa ttccagatcct ct

32

```

<210> ?
<211> 1602
<212> DNA
<213> Homo sapiens

<400> ?
atgaagatga gacgctacaa gcttttctc atgttctgtat tggccggcct gtgcctcatc
tccttcctgc acttcttcaa gaccctgtcc tatgtcaccc tcccccgaga actggcctcc
ctcagcccta acctggtgtc cagcttttc tggaaacaatg cccccgtcac gccccaggcc
agccccgagc caggaggccc tgacctgtcg cgtacccccc tctactccca ctcgcccctg
ctgcagccgc tgccgcccag caaggccggc gaggagctcc accgggtggaa cttggtgctg
cccgaggaca ccaccggata ttctgtgcgc accaaggcccg gcggcgtctg cttcaaaaccc
ggcaccacaaga tgcgtggagag gcccggcccg ggacggcccg aggagaagcc tgagggggcc
aacggctct cggccggcg gccaccggg tacctctgtat gcccgggaa ggcacggggg
ggccgaggcg cccggcgcaa gtgggtggag tgctgtgtcc tgccggctg gcacggacc
agctgcggcg tgcccactgt ggtgcagttac tcacaacctgc ccaccaagga gcggctgggt
cccgaggagg tgccgcccgg cgtcatcaac gcatcaac gcaaccacga gttcgacctg
ctggacgtgc gttccacgat gctggggcagc gtgggtggacg cttttgtgtt gtgcgagtcc
aacttcacgg ctatggggat gccggggcg ctcaagttcc gggagatgtt gaccatggc
accttcgagt acatccgcca caaggtgtcc tatgttcttc tggcacactt cccggccggc
ggccggcgagg acggctggat cgccgacgac tacctgtgcga ctttcctcac ccaggacggc
gtctcgccgc tgcccaacctt gcccggcgcac gacgttctca tcatttgcga tgccggacgag
ccctggccc gtgacggcgat ctttttctc aagttctacg atggctggac cgagcccttc
gccttcacaa tgcccaactt gctctacggc ttcttctggat agcaggccggg caccctggag
gtgggtgtcag gctgcacggc ggacatgtcg caggcagtgt atggctggat cggcatccgc
ctgcgcggcc gccaggacta caccatggcc aacttcacgat agtatagaaa ccgcaccggc
cacatccctgg tgccatgtggc gctggggcagc cccctgtcaact tcgcccgtcg gcactgtcc
tgggtgttca cggccggaggcatctacttc aagctgtgtt ccgcccggaa tggcgacttc
ccacgctgggg gtgactacgaa ggacaaggcg gacctgtactt acatccgcgg cttgtatccgc
accggggggctt ggttcgacgg cacggcagc gaggatccgcg ctgcagaccc cagcggacac
atgtatgcgc ccaaggatccctt gctgaagaac tacgaccggat tccactactt gttggacaaac
ccctaccagg agccggaggacacggccggc ggcgggtggc gccacaggggg tcccgagggaa
aggccggcccg cccqqqqqcaacttggacgat gctggaaagtctt ag

```

<210> 8
<211> ?027
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

```

<400> 8
catgattacg ccaagcttagc ggccgcattc ccgggaagct aggccaccgt ggcccgcctg
caggggaaagc ttgcatgcct gcagatcccc ggggatcctc tagagtgcac ctgcagtgca
gcgtgaccccg gtctgtcccc tctctagaga taatgagcat tgcatgtcta agttataaaa
aattaccaca tattttttt gtcacactt tttgaagtgc agtttatcta tctttataca
tatatttaaa cttaaatcta cgaataatat aatctatagtg actacaataa tattcgttt
tttagagaatc atataaatgatc acatggtagac atggctaa ggacaattga gtattttgac
aacaggactc tacagtttta tcttttagt gtgcgtgtt tttttttt ttttgcaat
agcttcacccat atataatact tcatttcattt tatttagtaca tccatttagg gtttagggg
aatggttttt atagactaat ttttttagtata catctatTTT attctatTTT agcctctaaa
ttaagaaaac taaaactcta ttttagttt ttatTTTaaat aatttagata taaaatagaa
taaaataaaag tgactaaaaaa taaaacaaaat accctttaag aaattaaaaaa aactaaggaa
acatTTTCT tttttcgagt agataatgcc accctgttAA acggcgtcga cggatctaac
ggacaccaac cagcgaacca gcagcgtcgc gtccggccaa gcaagcaga cggcacggca
tctctgtcgc tgccctgtggc cccctctcgat gagttccgcg ccaccgttgg acttgctccg
ctgtcgcgtat ccagaaatttgc cgtggccggag cggcagacgt gagccggcac ggcaggcggc
ctcctctcc ttcacggca cggcagctac gggggattcc ttcccacccg ctccttcgt
ttcccttcct cggccgcgt aataaaataga cacccccctcc acacccttcc tccccaaacct
cgttgtttc ggagcgcaca cacacacaac cagatctccc ccaaattccac ccgtcggcac
ctccgcctca aggtacgccc ctcgtccctt ccccccccccc ctctctacct tctctagatc
ggcgtttccgg tccatgcgtat gtttagggccc ggtagttcta ttctgttca tttttgttt
agatccgttgc ttgtttagatc tccgtgtcgc tagcgttgc acacggatgc gacctgtacg
tcagacacgt tctgattgtc aacttgcacg ttttctctt tggggatcc tgggatggct
ctagccgttc cgagacggg atcgattca tgatTTTTTG tggggatcc tgggatggct
qgttgcctt cttcccttat ttcaatataat ggcgtcact tggggatcc tgggatggct

```

catgttttt	tttgcttgg	ttgtgatgat	gtggctgg	tggcggtcg	ttctagatcg	1500
gagttagaaat	ctgttcaaaa	ctacctgggt	gattattaa	tttggatct	gtatgtgt	1560
gcacatacata	ttcatagtt	cgaattgaag	atgatggatg	gaaatatcg	tctaggatag	1620
gtatatacg	tgatcggtt	tttactgatg	catatacaga	gatgctttt	gttcgttgg	1680
ttgtgatgat	gtgggtgtt	ttggcggtcg	ttcattcg	ctagatcgga	gtagaatact	1740
gtttcaaact	acctgggt	tttattaatt	ttggaaactgt	atgtgtgt	catacatctt	1800
catagttacg	agtttaagat	ggatggaaat	atcgatctag	gataggtata	catgttgtat	1860
tgggtttac	tgatgcata	acatgatg	atatgcagca	tctattcata	tgctctaacc	1920
ttgagtagct	atctattat	ataaacaaat	atgtttata	attatttga	tcttgatata	1980
cttggatgat	ggcatatg	gcagctat	gtggat	ttagccctgc	cttcatacgc	2040
tatttattt	cttggact	tttctttt	cgatgtcac	cctgttgtt	ggtgtactt	2100
ctgcagggt	ccccccgggt	cgaccatgt	aaggggcagc	caccaccacc	accaccacat	2160
ggtcgcgtt	gtagaaaccc	caacccgt	aatcaaaaaa	ctgcacggcc	tgtggcatt	2220
cagtctggat	cgcgaaaact	gtggaaattg	tcagcgttgg	tggaaagcg	cgttacaaga	2280
aagccggca	attgtgtgc	caggcagtt	taacgatcg	ttcgcgcgt	cagatattcg	2340
taattatcg	ggcaacgt	gttatcagcg	cgaagtctt	ataccgaaag	gttgggcagg	2400
ccagcgat	gtgctgcgtt	tcgatgcgt	cactcattac	ggcaaagtgt	gggtcaataa	2460
tcaggaatgt	atggagc	aggcggtca	tacgcattt	gaagcgcgt	tcacgcgt	2520
tgttattg	ggggaaatgt	tactgtatc	cgttgtgt	aacaacgac	tgaactggca	2580
gactatccc	ccggaaatgg	tgattaccg	cgaaaacggc	aagaaaaaagc	agtcttactt	2640
ccatgat	ttaactat	ccggaaatc	tcgcagcg	atgctctaca	ccacgcgaa	2700
cacctgggt	gacgatatc	ccgtgggt	gcatgtcg	caagactgt	accacgcgt	2760
tgttga	cagggtgtt	ccaatgg	tgtcagcg	gaactgcgt	atgcggatc	2820
acagggtt	gcaactgg	aaggcact	cgggactt	caagtgg	atccgcac	2880
ctggcaacc	ggtgaagg	atctctat	actgtgcgt	acagccaaa	gccagacaga	2940
gtgtat	tacccgc	gcgtcggt	ccggctgt	gcagtgaagg	gcaacagtt	3000
cctgat	cacaaaccgt	tctacttac	tggcttgg	cgtcatgaa	atgcggact	3060
acgtgg	ggattcg	acgtgtgt	gtgcacg	cacgcattaa	tggactggat	3120
tggggcaac	tcctacc	cctcgat	cccttacg	gaagagatg	tcgactggc	3180
agatgaacat	ggcattgtt	tgattgat	aactgtgt	gtcggctt	acctctttt	3240
aggattgtt	ttcgaagc	gcaacaag	gaaagaact	tacagcga	aggcagtca	3300
cggggaaact	cagcaagc	acttacagg	gattaaag	ctgatagc	gtgacaaaaa	3360
ccaccca	gtgggtat	ggagtatt	caacgaa	gataccgc	cgcaagtgc	3420
cgggaaat	tcgcact	cggaagca	gcgttaa	gacccgc	gtccgatc	3480
ctgcgt	taatgtt	gcgcgct	caccgat	atcagcgt	tcttgatgt	3540
gtgtgc	aaccgtt	acggatgtt	tgtc	ggcatttgg	aaacggcaga	3600
gaaggta	gaaaaaag	ttctggct	gcaggag	ctgcatc	cgattatcat	3660
caccgaa	ggcgtgg	cgttagcc	gctgcact	atgtacacc	acatgtggag	3720
tgaagagt	cagtgt	ggctggat	gtatcacc	gtcttgc	gcgtcagc	3780
cgtcgt	gaacaggt	gaaatttgc	cgatttgc	acctcg	gcattatgc	3840
cgttgg	aacaagaa	ggtatctc	tcgcgacc	aaaccga	cggcgtt	3900
tctgctg	aaacg	ctggcat	cttgcgt	aaaccgc	agggaggca	3960
acaatgata	tgagtc	taaactg	gcactg	cgtctgt	gtctgattgt	4020
ttgtgat	ggtgg	tttggat	ttatc	tggtgtgt	tttatcatat	4080
gatcgat	ttgggtt	tttaacat	tgtaaaat	gtatctat	ataactcaat	4140
gtataagat	tgttcatt	tcgggtt	tagatct	tatttgac	gtgatgttt	4200
gactcc	acc	caactca	aactcat	atatgtcc	ctgttctt	4260
aagagtt	ctaccatt	agttgg	tatcgt	gcagcggc	tgtgtttt	4320
aacataac	ttgttac	atatat	cgccgg	agctagcc	ggtgccaga	4380
tccactag	ctagac	cgcttaat	actggcc	gttttac	gtcgact	4440
ggaaaaac	ggcgttac	aacttaat	ccttg	catcccc	tcgcagct	4500
gcgtat	gaagagg	gcaccat	cccttcc	catttt	gcctgaatgg	4560
cgaatgg	ctgatcg	atttct	tacgc	tgcgtt	cacaccgc	4620
atggtg	actc	tctgct	tgcc	ttaagcc	cccgcac	4680
gccaacac	gctgac	cctgac	ttgtct	ccggcatt	tttacagaca	4740
agctgt	gtctc	gctgc	tcagagg	tcaccgt	caccgaaac	4800
cgcgag	aagg	tgatc	tttgc	gttaat	tgataataat	4860
ggtttctt	acgtc	gcactt	ggaaat	cgcgg	ctattgtt	4920
attttctt	atacatt	atatgt	gtct	caataac	gataaaatgc	4980
tcaataat	tgaaaaa	agatgt	tattca	ttccgt	cccttattt	5040
ctttttcg	gcattt	ttcct	tgctcac	gaaacgt	tgaaagtaaa	5100
agatgct	gatc	tgca	gggtt	gaact	tcacacg	5160
taagatc	gagat	gccc	acgtt	atgat	cttttaaagt	5220
tctgctat	ggcgc	ttatcc	tgac	caagag	tcggtcgcc	5280
cataca	tctc	acttgg	gtact	gtcaca	gtcaca	5340
ggatgg	acgt	aattat	cgat	accat	agcatcttac	5400
ggccaa	cttct	cgat	tgct	ctaaacc	ataacactgc	5460
catggggat	catgt	cgctt	tggg	gagct	ttttgcacaa	5520
aaacgac	cgta	cgat	agca	aaacgt	aagccatacc	5580
aactgg	ctactt	tagt	ggca	gcaactatt	gcaactatt	5640
taaagg	tgca	ggacc	ttc	ggctgg	ttgctgataa	5700

actctggagcc	ggtgaagcgtg	ggtctcgccc	tatcattgca	gcactggggc	cagatggtaa	5760
gccctcccgt	atcgttagta	tctacacgac	ggggagtcag	gcaactatgg	atgaacgaaa	5820
tagacagatc	gctgagatag	gtgcctca	gattaagcat	tggtaactgt	cagaccaagt	5880
ttactcata	atactttaga	ttgatTTAA	acttcattt	taatttAAA	ggatcttagt	5940
gaagatcctt	tttgataatc	tcatgaccAA	aatccctAA	cgtgagTTT	cgttccactg	6000
agcgtcagac	cccgtagaaa	agatcaaagg	atcttcttga	gatcTTTT	ttctgcgcgt	6060
aatctgctgc	ttgcaaaacAA	aaaaaccacc	gctaccagcg	gtggTTT	tgccggatca	6120
agagctacca	actcttttcc	cgaaggtaac	tggcttcagc	agagcgcaga	taccaaatac	6180
tgtccttcta	gtgtagccgt	agttaggcca	ccacatcaag	aactctgtag	caccgcctac	6240
atacactcg	ctgctaattc	tgttaccatg	ggctgctgCC	agtggcgata	agtctgtct	6300
taccgggttg	gactcaagac	gatagttacc	ggataaggcg	cagcggcgtgg	gctgaacggg	6360
gggttcgtgc	acacagcccc	gcttggagcg	aacgacctac	accgaactga	gataacctaca	6420
gcgtgagcat	tgagaaagcg	ccacgcttcc	cgaagggaga	aaggcggaca	ggtatccgt	6480
aagcggcagg	gtcggAACAG	gagagcgcac	gagggagctt	ccagggggaa	acgcctggta	6540
tcttata	atgttccgggt	ttcgcCACCT	ctgacttgag	cgtcgattt	tgtgatgctc	6600
gtcagggggg	cggaggcttat	gaaaaaacgc	cagcaacgcg	gcctttttac	ggtcctggc	6660
cttttgcgtgg	ccttttgc	acatgttcc	tcctgcgtta	tcccctgatt	ctgtggataa	6720
ccgttattacc	gcctttgagt	gagctgatac	cgctcgcgc	agccgaacga	ccgagcgcag	6780
cgagtca	tgccggat	ggaaagagcg	cccaatacgc	aaaccgcctc	tccccgcgc	6840
ttggccgatt	cattatgtca	gctggcacga	caggttccc	gactggaaag	cggcagtga	6900
gccaacgcA	attaatgtga	gttagctcac	tcattaggca	ccccaggctt	tacactttat	6960
gcttccggct	cgtatgttgt	gtggaatttgt	gagcggataa	caatttcaca	caggaaacag	7020
ctatgac						7027

<210> 9
<211> 6818
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 9						
cctgcagatc	cccggggatc	ctctagagtc	gacctgcagt	gcagcgtgac	ccggcgtgtc	60
ccctctctag	agataatgag	cattgcatgt	ctaagttata	aaaaattacc	acatattttt	120
tttgtcacac	ttgtttaa	tgcatgttat	ctatctttat	acatataattt	aaactttaat	180
ctacgaataa	tataatctat	agtactacaa	taatatcagt	gttttagaga	atcatataaa	240
tgaacagtta	gacatggct	aaaggacaat	tgagttttt	gacaacagga	ctctacagtt	300
ttatctttt	agtgtgcatg	tgttctcc	tttttttgc	aatagctca	cctatataat	360
acttcatcca	tttttattatg	acatccattt	agggtttagg	gttaatgtt	tttataagact	420
aattttttt	gtacatctat	tttattctt	tttgcctc	aaattaaaga	aactaaaact	480
ctatTTTT	ttttttattat	aataattttat	atataaaata	gaataaaaata	aagtgactaa	540
aaattaaaca	aataccctt	aagaaattaa	aaaaactaag	aaaacatTTT	tctgtttcg	600
agtagataat	gccagcctgt	taaacgcctg	cgacgagct	aacggacacc	aaccagcga	660
ccagcagcgt	cgcgtgggc	caagcgaagc	agacggcag	gcatctctgt	cgctcctct	720
ggacccctct	cgagagtcc	gttccaccgt	tggacttgct	ccgctgtcg	catccagaaa	780
ttgcgtggcg	gagcggcaga	cgtaggcgg	cacggcaggc	ggcctccctc	tcctctcag	840
gcacggcagc	tacggggat	tccttccca	ccgcctcc	gctttccctt	cctcgcggc	900
cgtataataat	agacaccccc	tccacaccc	ctttcccca	cctcgttgg	tccggagcgc	960
acacacacac	aaccagatct	cccccaatc	caccgcgtgg	cacccgcgt	tcaaggatcg	1020
ccgctgtcc	cccccccccc	ccccctctca	ccttctctag	atcggcgttc	cggtccatgc	1080
atggtttaggg	cccggttagtt	ctacttctgt	tcatgtttgt	gttagatccg	tgtttgtgtt	1140
agatccgtgc	tgctagcg	cgtacacgg	tgcgacctgt	acgtcagaca	cgttctgatt	1200
gctaacttgc	cagtgtttc	tttggggaa	tcctggatg	gctctagccg	ttccgcagac	1260
gggatcgatt	tcatgtatTTT	tttgcattc	ttgcataggg	tttgcattc	cctttccctt	1320
tatttcaata	tatggcgtgc	acttgcTTT	cgggtcatct	tttgcattc	ttttttgtct	1380
ttgttgtgtat	gtatgggtct	gttggggcg	tcgttctaga	tcggagtaga	attctgtttc	1440
aaactacctg	gtggatttat	taatttttgg	tctgtatgt	tgtgcattac	atattcatag	1500
ttacgaattt	aagatgtatgg	atggaaatat	cgatcttagg	taggtatata	tgttgatgcg	1560
ggttttactg	atgcataatac	agagatgctt	tttgcgtc	ttgttgatgt	gatgtgggt	1620
ggttggggcg	tcgttcatc	gttctagatc	ggagtagaa	actgtttcaa	actacctgtt	1680
gtatTTTT	atTTTGAAC	tgtatgtgt	tgtcatacat	tttgcattt	acgagtttac	1740
gatggatgga	aatatcgatc	taggataggt	atacatgttg	atgtgggtt	tactgtatgc	1800
tatacatgtat	ggcatatgca	gcatttattt	atatgtctca	accttgcgt	cctatctatt	1860
ataataaaca	agtatgtttt	ataattttt	tgatcttgc	ttgttgatgt	gatggcatat	1920
gcagcagcta	tatgtggatt	tttttagccc	tcgccttc	cgcttatttt	tgtcttggta	1980
ctgtttctt	tgtcgatgt	caccctgttg	tttgcgttta	tttgcgttgc	gtaccccccgg	2040
ggtcgaccat	ggtgtatgaga	cgctacaagc	tcttctcat	tttgcgtatg	gcccggcctgt	2100
gcctcatctc	cttcctgcac	tttcaaga	ccctgtcc	tgtcaccc	ccccgagaac	2160
tggccctccct	cagcccta	ctgggttcca	gtttttctg	gaacaatgcc	ccggcgtacgc	2220

cccaaggccag	ccccgagcca	ggaggcccgt	acctgctgcg	taccccactc	tactcccact	2280
cgccttgcgt	gcagccgctg	ccgcccagca	aggcgccgca	ggagctccac	cgggtggact	2340
tggtgtgcc	cgaggacacc	accgagtatt	tcgtgcgcac	caaggccggc	ggcgtctgct	2400
tcaaaacccgg	caccaagatg	ctggagaggc	cccccccccgg	acggccggag	gagaaggcctg	2460
agggggccaa	cggcttctcg	gccccggcggc	caccccggtt	cctccctgagc	gcccgggagc	2520
gcacgggggg	ccgaggcgcc	ccgcgcaagt	gggtggagtg	cgtgtccctg	cccggtctggc	2580
acggacccag	ctgcggcgtg	cccactgtgg	tgcagtaactc	caacctgccc	accaaggagc	2640
ggctgggtcc	cagggaggtg	ccgcgcggcg	tcataaacgc	catcaacgtc	aaccacgagt	2700
tcgacctgtc	ggacgtgcgc	ttccacagac	tgggcacgt	ggtggacgcc	tttgggtgt	2760
gcgagttcca	cttcacggct	tatggggag	cgcggccgt	caagttccgg	gagatgctga	2820
ccaatggcac	cttcgagtag	atccgcaca	aggtgctcta	tgtttctcg	gaccacttcc	2880
cggccggcgg	ccggcaggac	ggctggatcg	ccgacgacta	cctgcgcacc	ttcttcaccc	2940
aggacggcg	ctcgccgctg	cgcaaacatgc	ggcccgacga	cgttttcatc	attgacgatg	3000
cgagcagat	ccggggccgt	gacggcggtc	ttttctcaa	gctctacat	ggctggaccg	3060
agcccttcgc	cttccacatg	cgcaagtgc	tctacggctt	tttctggaa	cagccgggca	3120
ccctggaggt	gggtgtcaggc	tgcacgggtt	acatgctgca	ggcagtgtat	gggtggacg	3180
gcatccgcct	gcgcgcgcgc	cagtactaca	ccatggccaa	tttcagacag	tatgagaacc	3240
gcacccggcc	catccgtgt	cagtggcgc	tggcagcccc	cctgcactt	gcccgtggc	3300
actgtctctg	gtgcttcac	cccgaggggc	tctacttcaa	gctcgtgtcc	gcccagaatg	3360
gcaacttccc	acgctgggggt	gactacgagg	acaaggggg	cctgaactac	atccgcggcc	3420
tgatccgcac	cgggggctgg	tgcacggca	cgacgaggaa	gtacccggct	gcagacccca	3480
gcaagcacat	gtatgcgc	aagtacatgc	tgaagaacta	cgaccggg	cactacctgc	3540
tggacaaccc	ctaccaggag	cccaggagca	cgccggcggg	cggtggcgc	cacagggg	3600
ccgagggaa	gcccggcc	cgggggcaaa	tggacgaggc	ggaagtgc	aaaaactca	3660
tctcagaaga	ggatctgaat	taggatcc	ggtttaact	gagggcactg	aagtgcctt	3720
atgtgtcga	ttgttgtgt	ttgttgtgt	gtattttgtt	taataata	agcatggct	3780
tgattttatc	atatgatcga	tctttgggtt	tttttttttt	acattgtaaa	atgtgttatct	3840
attaataact	caatgtat	gtatgttca	ttcttcgggt	gcatagatc	tgcttatttt	3900
acctgtat	ttttgtactcc	aaaaacaaa	atcacaactc	aataaactca	tggaaatatgt	3960
ccacctgttt	tttgcagag	tcatctacca	ttccagttgg	catttatcag	tgtgcagcg	4020
gcgcgtgtct	ttgttaacata	acaattgtt	acggcatata	tccacggcc	gcctagctag	4080
ccacgggtgc	cagatccact	agttctagag	cgcccgctt	attcacttgc	cgtcg	4140
caacgtcg	actggggaaa	ccctggcg	acccaaactt	atcgcttgc	agcacatccc	4200
ccttcgcca	gtggcgtaa	tagcgaagag	gcccgcac	atcgccctt	ccaaacagtt	4260
cgacgcgt	atggcgaat	gcgcgtat	cggtatttt	tccttac	tctgtgcgg	4320
atttcacacc	gcatatgtgt	cactctcgt	acaatctgt	ctgatgc	atagttaa	4380
cagccccgac	acccgccaac	acccgctgac	gcccctgac	gggcttgc	gtctccggc	4440
tccgcttaca	gacaagctgt	gaccgtctt	gggagctgc	tgtgtcagag	gttttcc	4500
tcatcaccga	aaacgcgcg	acgaaagg	ctcgtgata	gcctattttt	ataggtta	4560
gtcatgataa	taatggttt	ttagacgtca	ggtggcactt	ttcggggaaa	tgtgcgcgg	4620
acccctattt	gttttatttt	ctaaatacat	tcaaataatgt	atccgc	gagacaataa	4680
ccctgataaa	tgcttcaata	atattgaaa	aggaagagta	tgatatttca	acatttccgt	4740
gtgccttca	ttccctttt	tgccggat	ttcccttct	tttttgc	cccaagaaacg	4800
ctgggtgaa	taaaaagatgc	tgaagatc	ttgggtgcac	gagtgggtt	catcgaactg	4860
gatctcaaca	gcggtaagat	ccttgagagt	tttcgcggc	aagaacgtt	tccatgtat	4920
agcactttt	aagtctgt	atgtggcg	gtattatccc	gtattgac	cgggcaagag	4980
caactcggtc	gccgcata	ctattctc	aatgacttgg	ttgagtactc	accagtac	5040
aaaaagcata	ttacggatgg	catgacat	agagaattat	gcagtgc	cataaccat	5100
agtataaca	ctgcggccaa	cttacttct	acaacatgc	gaggaccgaa	ggagctaacc	5160
gttttttgc	acaacatggg	ggatcat	actgcctt	atcg	accggagctg	5220
aatgaagca	taccaacg	cgacgt	accacatgc	ctgttag	ggcaacaacg	5280
ttgcgc	tttttttt	cgactactt	actctagtt	ccggcaaca	attaatagac	5340
tggatggagg	cggataaa	tgcaagg	tttgcgt	cgcccttcc	ggctggctgg	5400
tttattgtc	atcaaactt	agccgtt	cgtgggtt	cggttat	tgcagactg	5460
gggcccggat	gtaa	ccgtatcg	tttccgaa	cgacgggg	tcaggcaact	5520
atggatga	aaaatagaca	gatcg	atagg	cactgat	gcattggtaa	5580
ctgtcagacc	aagttaactc	atataat	tttttttt	taaaactca	ttttttaattt	5640
aaaaggat	aggtga	ttttttt	tttttttt	ccaaatccc	ttaacgtgag	5700
ttttcg	actgacgt	agacccc	gaaaatgc	aaggat	tttgc	5760
tttttgc	gcgtat	tttttttt	acaaaaaa	caccg	tttgc	5820
tgtttcc	atcaagag	accaactt	tttccgaa	taactgg	cagcgtt	5880
cagat	atactgt	tctatgt	ccgtat	ccacc	tttgc	5940
gtacacc	ctacata	tttccgaa	atccgtt	cgatgg	tttgc	6000
gataagt	gtcttac	tttccgaa	tttgcgt	taccgg	tttgc	6060
tcgggct	cggggg	tttccgaa	tttgcgt	tttgc	tttgc	6120
ctgagat	tacag	tttccgaa	tttgcgt	tttgc	tttgc	6180
gacaggtat	cggtaa	tttccgaa	tttgcgt	tttgc	tttgc	6240
gaaacgcct	ggat	tttccgaa	tttgcgt	tttgc	tttgc	6300
tttttgc	gtctgt	tttccgaa	tttgcgt	tttgc	tttgc	6360
ttacgg	tggc	tttccgaa	tttgcgt	tttgc	tttgc	6420
gattctgt	ataacc	tttccgaa	tttgcgt	tttgc	tttgc	6480

acgaccggagc	gcagcgagtc	agttagcgag	gaagcggaaag	agcgcccaat	acgcaaaccg	6540
cctctccccg	cgcgtggcc	gattcattaa	tgcagctggc	acgacagggtt	tcccactgg	6600
aaagcgggca	gtgagcgaa	cgcaattaat	gtgagtttagc	tcactcatta	ggcaccccgag	6660
gctttacact	ttatgttcc	ggctcgatg	tttgttgaa	tttgtgagcgg	ataacaattt	6720
cacacaggaa	acagctatga	ccatgattac	gccaaagctag	cggccgcatt	cccgggaagc	6780
taggccaccg	tggccgcct	gcaggggaag	cttgcatt			6818

<210> 10
<211> 7545
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 10						
cgattaaaaa	tctcaattat	atttggctta	attttagttt	gtatttagtta	aaacaaattt	60
gaaccaaacc	aaaatataaa	tatatagttt	ttatataatat	gccttaaga	ctttttatag	120
aatataatctt	aaaaaatatc	tagaaattat	tgccactt	ctggcatgtt	atatttcgtt	180
aaatataatgg	tgctccattt	ttatataactt	taaataattt	gttgtacgt	cactttctta	240
tcaagtgtta	ctaaaatgcg	tcaatcttt	tgttcttcca	tattcatatg	tcaaaaccta	300
tcaaaaattct	tatataatctt	tttgcattt	gaagtggaaat	ttcgataatt	taaaatttttt	360
tagaacatata	cattatttat	gtatcatatt	gattttata	cttaattact	aaatttggtt	420
aactttgaaa	gtgtacatca	acgaaaaattt	agtccaaacga	ctaaaataaa	taaatatcat	480
gtgttattaa	gaaaattctc	ctataagaat	attttaatag	atcatatgtt	tgtaaaaaaaaa	540
attaattttt	actaacacat	atattttactt	atcaaaaattt	tgacaaaatg	agataaaaat	600
aatatttcatt	taacaaaaaa	aaaaccaggaa	aatgtggaa	acccggccaa	accgaaccaa	660
tccaaaccg	tatagggtgt	ttgggtttagt	tttgatataa	accgaaccaa	ctcggtccat	720
ttgcacccct	aatcataata	gctttatata	ttcaagatata	tattaagttt	acgttgtcaa	780
tatcctggaa	attttgcaaa	atgaatcaag	ccttatatggc	tgtatataatg	atttaaaagc	840
agctcgatgt	ggtggtaata	tgtatatttt	tttgattctaa	aaaatataatc	caagtattaa	900
taatttctgc	taggaagaag	gttagctacg	atttacagca	aagccagaat	acaatgaacc	960
ataaaagtgt	tgaagctcg	aatatacgaa	ggAACAAATA	ttttttaaaaa	aatacgcaat	1020
gacttggaaac	aaaagaaaatg	gatataatttt	ttgttcttta	acaagcatcc	cctctaaaga	1080
atggcagggtt	tcctttgcatt	tttgcatttt	tgctcccttc	gttacaaaaaa	tttggacta	1140
ctattttggaa	cttcttctga	aaatagtggc	caccgcttaa	ttaaggcgcg	ccatgcccgg	1200
gcaaggccg	gcttaattaa	atttttatgt	ttaaacttagg	aaatccaagc	ttgggctgca	1260
ggtaatccc	attgtttttt	aaggcgttca	acattgtatc	cttctcgag	gtcattcata	1320
tgcttgagaa	gagatcgccc	atagtccaaa	ataaaacaaa	ggtaagat	cctggtaaaa	1380
agtggaaaaca	tcaatggaaa	ggtggatataa	agtaaaatat	cgtaataaa	aggggccca	1440
aagtggaaaat	tactttttt	tacttatttt	aaaatttgggg	atgtttttgt	cggactttt	1500
atacgtcatt	tttggatgaa	tttgggtttttt	agtttatttt	ctttttggaaa	tgcataatctg	1560
tatgttggatc	gggttttaag	ttcggtttgtt	tttttttttt	cagaggggat	ttgtataagaa	1620
atatctttaa	aaaaacccat	atgtcaattt	gacataattt	tttgagaaaaa	tatataattca	1680
ggcgaattct	cacaatgaac	aataataaga	ttaaaaatagc	tttccccctgt	tgcagcgcatt	1740
gggttattttt	tctatggaaa	ataaaagata	aacttagact	aaaaacattt	acaaaaacaa	1800
ccccctaaagt	tcctaaagcc	caaagtgcata	tccacgatcc	atagcaagcc	cagcccccaacc	1860
caaccccaacc	caacccaccc	cagtcaccc	aactggacaa	tagtctccac	accccccac	1920
tatcaccgtt	agttgtccgc	acgcacccca	cgttcgcac	ccaaaaaaa	aaaaagaaaag	1980
aaaaaaaaaa	aaaagaaaaaa	acagcagggt	gttcgggtc	gtggggggcg	gaaacgcgag	2040
gaggatcgcc	agccacgcg	gaggccggcc	ctccctccgc	ttccaaagaa	acgccccccca	2100
tcgcccactat	atacatacc	ccccctctcc	tcccatcccc	ccaaaccttac	caccaccacc	2160
accaccaccc	ccaccccttc	ccccctcgct	gccggacgac	gcctcccccc	tccccctccg	2220
ccgcccggcc	gccggtaacc	accccgcccc	tctcctttt	cttctccgt	tttttttttc	2280
cgtctcggtc	tcgatcttt	gccttggtag	tttgggtggg	cgagaggcgg	tttcgtgcgc	2340
ccccagatcg	gtgcgcggg	ggggcgggat	ctcgcggctg	gggctctcgc	cggcgtggat	2400
ccggcccccgg	tctcgccgggg	aatggggctc	tcggatgttag	atctgcgtac	cggcgttttt	2460
ggggggagat	atgggggggtt	taaaatttttcc	gcccatttttt	acaagatgt	gaagagggga	2520
aaaggggcact	atggggggat	ttttttttttt	tttcgtctgt	ttcgtcaggc	tttagatgtgc	2580
tagatctttt	tttcttcttt	tttgggggtt	tttttttttt	cctcagcatt	tttcatcggt	2640
agtttttttt	tttcatgtttt	tttgggggtt	tttttttttt	cggagctttt	ttttaggttag	2700
accatggctt	ctccggagag	gagaccagg	gagatttagc	cagctacagc	agctgtatgt	2760
gccgcgggtt	gtgatatcg	taaccattac	attgagacgt	ctacagtggaa	ctttaggaca	2820
gagccacaaa	caccacaaga	gtggattgtt	gatcttagaga	gtttgcaga	tagataccct	2880
tggttgggtt	ctgggggtt	gggtgtttgt	gctgggtattt	cttacgttgg	gccttggaaag	2940
gcttaggaacg	cttacgttgg	gacagtgttt	agttactgttt	acgtgtcaca	taggtatcaa	3000
aggttggggcc	tagatccac	attgtacaca	catttgcattt	agtctatgg	ggcgaaggt	3060
ttaatgtctg	tggttgtctgt	tataggccctt	ccaaacgtatc	catctgttag	tttgcatttag	3120
gctttggat	acacagcccg	gggtacattt	cgcgcagctg	gatacaagca	tttggatgg	3180
catgtgttgc	gtttttggca	aagggtttttt	gatgttgcag	ctccctccaa	gccagttagg	3240

ccagttaccc	agatctgagg	taccctgagc	tcggcgcag	cgtgtcggt	tccgtcgta	3300
gttctggccg	gccgggcctt	gggcgcgcga	tcagaagcgt	tgcgttgccg	tgtgtgtgc	3360
tctggtttgc	ttaattttta	ccaagtttgc	ttcaagggtgg	atcgcttgtt	caaggcccgt	3420
gtgctttaaa	gaccacccgg	cactggcagt	gagtgtgttgc	gcttgtgttag	gctttggtag	3480
gtatgggctt	tatttgc	ttggatgttgc	gtactacttgc	ggtttggta	attattatgaa	3540
gcagttgcgt	attgtattc	agctggcgtt	cctggacattt	gttatgtattt	aataaatgct	3600
ttgctttctt	ctaaagatct	ttaagtgcgt	aattcatattt	tcctcctgca	gggtttaaac	3660
ttgcccgtggc	ctatttcag	aagaagttcc	caatagtagt	ccaaaattttt	tgtaacgaag	3720
ggagcataat	agttacatgc	aaaggaaaaac	tgccatttctt	tagaggggat	gcttggtaaa	3780
gaacaaaaaa	tatatactt	tctttgttc	caagtcatttgc	cgtatttttt	taaaaatattt	3840
tgttccttcg	tatatttcga	gcttcaatca	ctttatgttt	ctttgttattt	tggcttgc	3900
gttaaatctgt	gctaaccctt	ttccttagcag	aaattttaaa	tacttgggat	attttttag	3960
aatcaagtaa	attacatattt	accaccatc	cgagctgttt	ttaaatttcat	attacagcca	4020
tataggcttg	attcatttttgc	aaaattttcc	aggatatttgc	caacgttaac	ttaataatat	4080
cttggaaatat	taaagctattt	atgatttaggg	gtgcaaatttt	accgagttgg	ttcgggtttat	4140
atcaaaatca	aaccaaaacca	actatatcg	tttggttgg	ttcgggttttgc	ccgggttttc	4200
agcattttctt	ggtttttttt	ttgttagatg	aatattttttt	taatcttact	ttgtcaaaattt	4260
tttgataagt	aaatatatgt	tttagttaaaa	attaattttttt	tttacaaaca	tatgtatctat	4320
taaaaatattt	ttataggaga	attttcttaa	taacacatga	tattttttttt	tttttagtgc	4380
ttgactaattt	tttcgttgc	gtacactttc	aaaggtaacc	aaatttttagt	attaagtata	4440
aaaatcaata	tgatacttaa	ataatgtat	gttctattttt	attttaaattt	atcgaaattt	4500
cacttcaaat	tcgaaaaaaa	tatataagaa	ttttgtatga	ttttgacata	tgaatatggaa	4560
agaacaaaaga	gattgacgca	tttttagtaac	acttgataag	aaagtgtatc	tacaaccaat	4620
tattnaaagt	taataaaaat	ggagcacttgc	atatttaacg	aaatatttaca	tgccagaaga	4680
gtcgcaaata	tttcttagata	tttttttaaag	aaaattctat	aaaaggctt	aaaggcatat	4740
atataaaaac	tatataatttgc	tatttttgttgc	ttgttgcgtt	ttgttttact	caataccaaa	4800
ctaaatttgc	ccaaatataa	ttgggatttttgc	taatcgcggc	ccactagtca	ccgggttagc	4860
ttggcgtat	catggcata	gttgcgttgc	gtgttgcatttgc	tttatccgc	cacaatttcca	4920
cacaacatac	gagccggaaag	cataaaatgttgc	aaaggcttggg	gtgcctaatttgc	agttagctaa	4980
ctcacattaa	ttgcgttgc	ctcactgc	gttgcgttgc	cgggaaacctt	gtcggtccag	5040
ctgcattat	gaatcgcc	acgcgcgggg	agaggcggttgc	tgcgatattttgc	gcgtctttcc	5100
gctgcgcac	ctgcgcac	tgcgacgc	tcctcgctca	ctgactcg	gcgtcggc	5160
gttcggctgc	ggcgagcgg	atcagcttgc	tcaaaggcgg	taatacgttgc	atccacagaa	5220
tcaggggata	acgcaggaaa	gaacatgttgc	gcaaaaggcc	agcaaaaggc	cagaaccgt	5280
aaaaggccg	cggtgcttgc	ttttttccat	aggctccgc	ccccctgac	gcatcacaaa	5340
aatcgacgt	caagttagag	gtggcggaaac	ccgcacaggac	tataaagatgc	ccaggcg	5400
ccccctggaa	gctccctcg	ggcgtcttgc	gttccgaccc	tgccgcttac	cgatcacctg	5460
tccgccttgc	tccctcg	aagcgttgc	ctttctcata	gtcgttgc	taggtatctc	5520
agttcgttgt	aggtcg	cttcaagcttgc	ggctgtgtgc	acgaaccccc	cggtcagccc	5580
gaccgctgc	ccttatccgg	taactatcg	ctttagtcca	acccggta	acacgactta	5640
tcgcaactgg	cagcagcc	ttgttacagg	attagcagag	cgaggatgt	aggcggtgt	5700
acagagtttgc	tgaagtgg	gccttactac	ggcttactac	gaaggacgt	atttggat	5760
tgcgtctgc	tgaaggcc	taccttgc	aaaaggatgttgc	gtatcttgc	atccggcaaa	5820
caaaccac	ctggtagc	tttttttttgc	ttttgttgc	acgcgttac	gcgcagaaaa	5880
aaaggatctc	aaagaatgttgc	tttgcatttttgc	tctacgggttgc	ctgacgttgc	gttgcacgaa	5940
aactcacttgc	aagggttttgc	ggtcatgaga	ttatcaaaaa	ggatcttgc	ctagatcctt	6000
ttaaaatattt	aatgaatgttgc	taatcaatc	taaagtat	atgatgttgc	ttgttctgac	6060
agttaccaat	gtttaatcg	tgaggcac	atctcgttgc	tctgttctt	tcgttcatcc	6120
atagttgc	gactcccg	cggttgc	actacgttgc	gggggggtt	accatctggc	6180
cccagtgc	caatgtatcc	gctgaccc	cgcttcc	ctccagattt	atcagcaata	6240
aaccagcc	ccggaaagg	cgagcgc	agtgttgc	caacttac	cgctccatc	6300
cagtcttgc	atttgc	ggaagcttgc	ttttgttgc	cgccagtttac	tagttgc	6360
aacgttgc	ccatttgc	aggcatgttgc	gtgttgc	cggttgc	tatggcttca	6420
ttcagcttgc	gttccaa	atcaaggc	tttacatgttgc	ccccatgtt	gtcaaaaaaa	6480
gccccgttgc	ccttcgttgc	tccgtatgttgc	gtcagaagta	agttggcc	agtgttatca	6540
ctcatgttgc	ttggc	gcataatttgc	tttactgttgc	tgccatccgt	aatgttgc	6600
tctgttgc	gtgttgc	aaccaagt	ttctgttgc	agtgtatgc	gcgaccg	6660
tgctcttgc	ccggcgttgc	acgggataat	accgcgc	atagcgttac	ttttaaatgt	6720
ctcatcat	gaaaaacgttgc	ttccccggc	aaatcttgc	ggatcttacc	gttgc	6780
tccagttgc	tgttaccc	tcgtgc	aactgttgc	cagcatctt	tatccatc	6840
agcgttgc	ggttgc	aaacagg	aaaaatgc	caaaaaagg	aataaggc	6900
acacggaaat	gttgaatact	catacttgc	cttttcaat	attatttgc	catttatcg	6960
ggttatttgc	tcatgttgc	tttgc	aaatgttgc	agaaaaataa	acaaatagg	7020
gttccgcgc	catttcccc	aaaatgttgc	cctgtacgttgc	aagaaaccat	tattatcg	7080
acattaacct	ataaaaatag	gcgttac	aggcc	gtctcg	tttcgggt	7140
gacgggtgaaa	accttgc	catgc	ccggagacgg	tcacagcttgc	tctgttgc	7200
gatggccgg	gcagaca	ccgtc	ccgtc	gtgttgc	gtgtcgg	7260
ttggcttact	atgcgc	catgc	agagc	gtactgttgc	tgccatcat	7320
ataccgc	atgcgt	aaagc	aaatgc	cgcatc	ccatcg	7380
cgcaacttgc	ggttgc	gatgc	tttttgc	gcca	attcagg	7440
gggggatgttgc	ctgcaaggc	at	tttttgc	gttgc	gtc	7500

tgtaaaacga cggccagtga attacaccgg tgtgatcatg ggccg

7545

<210> 11
<211> 11643
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 11		
cgattaaaaa cccaattata tttggtctaa tttagtttg tattgagtaa aacaattcg	60	
aaccaaaacca aaatataaat atatagttt tatatatatg cctttaagac ttttataga	120	
attttcttta aaaaatatct agaaatattt ggcactcttc tggcatgtaa tattcgtta	180	
aatatgaagt gctccatttt tattaactt aaataattgg ttgtacgatc actttcttat	240	
caagtgttac taaaatgcgt caatctttt gtttccat attcatatgt caaaatctat	300	
caaattctt atatatctt ttcgatttt aagtggaaat tcgataattt aaaattaaat	360	
agaacatattt attatttagg tatacatatg attttatac ttaattacta aatttggta	420	
actttggaaag tgtacatcaa cgaaaaatata gtcaaacgac taaaataat aatatcatg	480	
tgttattaaag aaaattctcc tataagaata ttttaataga tcatatgtt gtaaaaaaaaa	540	
ttaattttta ctaacacata tatttactta tcaaaaattt gacaaagtaa gattaaaata	600	
atattcatct aacaaaaaaaaa aaaccagaaaa atgctgaaaa cccggcaaa ccgaaccaat	660	
cggggccat atagttggtt tggtttgatt ttgatataaa ccgaaccaac tcggtccatt	720	
tgcaccccta atcataatag cttaatatt tcaagatatt attaagttaa cgttgtcaat	780	
atccctggaaa ttttgcaaaa tgaatcaagc ctatatggct gtaatatgaa tttaaaagca	840	
gctcgatgtg gtggatatt gtaattttact tgattctaa aaaaatatccc aagttataat	900	
aattttctgt aggaagaagg ttagctacga tttacagcaa agccagaata caaagaacca	960	
taaagtgatt gaagctgaa atatacgaag gaacaaatat ttttaaaaaa atacgcaatg	1020	
acttggaaaca aaagaaaatgt atatattttt ttttcttaaa caagcatccc ctctaaagaa	1080	
tggcagttt ccttgcatt taactattat gtccttgcg ttacaaaaat ttggactac	1140	
tattggaaat tcttcgtaaa atagtggcca cgccttaatt aaggcgcgc atgccccctg	1200	
cagatcccg gggatccctc agagtcgacc tgcagtcagc cgtgaccgg tcgtgcccc	1260	
ctctagagat aatgagcatt gcatgtctaa gttaaaaaaa attaccacat atttttttg	1320	
tcacacttgt ttgaatgcg ttttatctt ctttatatac atatttaaac ttatctac	1380	
gaataatata atctatagta ctacaataat atcagtgttt tagagaatca tataaatgaa	1440	
cagttagaca tggcttaaag gacaatttag ttttgaca acaggactt acagttttat	1500	
cttttttagt tgcatgtgtt ctccttttt tttgcaaata gcttcaccta tataataactt	1560	
catccattttt attagtacat ccatttaggg ttttaggtta atggttttt tagactaatt	1620	
tttttagtac atctatttt ttctatTTTt gcctcttaat taagaaaact aaaactctat	1680	
tttagttttt ttatTTAata atttagataaaataaaatgaaat aaaataaagt gactaaaaat	1740	
taaacaataa ccctttaaga attaaaaaaa actaaggaaa cattttttt gttcgagta	1800	
gataatgcga gcctgttaaa cgccgtcgcg gactctaacc gacaccaacc agcgaaccag	1860	
cagcgtcgcg tcggggcaag cgaagcagac ggcacggcat ctctgtcgct gcctctggac	1920	
ccctctcgag agttccgctc caccgttggc cttgtccgc tgcggcatc cagaatttgc	1980	
gtggcggagc ggcagacgtg agccggcacg gcaggcggcc tcctccttctt ctacggcac	2040	
ggcagctacg ggggattcct tccccaccgc tccttcgctt ccccttc cccggccgt	2100	
ataaatagac accccctcca caccctctt ccccaaccc tcgttgcgc gaggcgcacac	2160	
acacacaaacc agatctcccc caaatccacc cgtcgacc tccgcttcaa ggtacgcccgc	2220	
tcgtcctccc cccccccccc tcttaccc tcttagatcg gcgttccggt ccatgcatgg	2280	
ttagggcccg gttagttctac ttctgttcat gtttggatgtt gatccgtgtt tggtttagat	2340	
ccgtgctgtc agcgttgcgt cacggatgcg acctgtacgt cagacacgtt ctgattgtca	2400	
acttgcctgtt gtttctctt gggaaatcct gggatggctc tagccgttcc gcagacggga	2460	
tcgatTTcat gatTTTTTTt gttcgttgc atagggtttt gtttgcctt ttccTTatt	2520	
tcaatataatg ccgtgcactt gtttgcggg tcatctttt atgcttttt ttgtcttgg	2580	
tgtgatgtatg tggcttgcgtt gggcggcgt tctagatcg tctatgttcc gatagaattt	2640	
tacctgggtt atttataat tttggatgtg tatgtgtgtt ccatacatat tcatatgttac	2700	
gaattgaaga ttagggatgg aaatatcgat cttagatagg tatacatatgtt gatgcgggtt	2760	
ttactgtatgc atatacagag atgttttttgc ttcgcttgcgt tgtagatgt tggtgtgg	2820	
gggcggcgtc tcattcggtt tagatcgag tagaataactg tttcaaaacta cctgggtgtat	2880	
ttattaaattt tggaaactgtt tggatgtgtc atacatcttcc atagttacga gtttaagatg	2940	
gatggaaata tcgatctagg ataggatatac atgttgcgtt gggttttact gatgcataata	3000	
catgatggca tatgcagcat ctattcatat gctctaacc ttagatccca tctattataaa	3060	
taaacaagta tggatTTTaa ttatTTTgtt ctgtatatac ttggatgtatg gcatatgcag	3120	
cagctatatg tggatTTTTt tagccctgc ttcatacgat atttatttgc ttgtactgt	3180	
ttcttttgc gatgtcacc ctgttgcgtt gtttacttc tgcagggtac cccgggggtc	3240	
gaccatgggtt atggacgtt acaagcttt tctcatgtt tgcggccgc gcctgtgcct	3300	
catctccctc ctgcattctc tcaagaccct gtcctatgtc acctcccccc gagaactggc	3360	
ctccctcagc ctaacctgg tggatgtgtt ttttgcggaaatggccggg tcacgccccca	3420	
ggccagcccc gagccaggag gcccgtaccc gctgcgtacc ccactctact cccactcgcc	3480	
cctgctgcagc ccgctgcccgc ccagcaaggc ggccgaggag ctccaccggg tggacttgg	3540	

gctgcccggag	gacaccaccg	agtatttctgt	gcgcaccaag	gccggcgccg	tctgcttcaa	3600
accgggcacc	aagatgctgg	agaggcccc	cccgggacgg	ccggaggaga	agcctgaggg	3660
ggggggacgg	tcctcgcccc	ggcggccacc	ccggtaacctc	ctgagcgc	gggagcgcac	3720
ggggggccgg	ggcgcggcgg	gcaagtgggt	ggagtgcgtg	tgccctgccc	gctggcacgg	3780
acccagctgc	ggcgtgcccc	ctgtgggtgca	gtactccaac	ctgcccacca	aggagcggct	3840
ggtgcccagg	gagggtggcgc	gccgcgtcat	caacgcacatc	aacgtcaacc	acgagttcga	3900
cctgctggac	gtgcgtttcc	acgagctggg	cgacgtggtg	gacgccttgc	tggtgtgcga	3960
gtccaacttc	acggcttatg	gggagccgcg	gccgctcaag	ttccgggaga	tgctgaccaa	4020
tggcaccttc	gagtagatcc	gccacaaggt	gtctatgtc	ttccctggacc	acttccgc	4080
cggcggccgg	caggacggct	gatgcggca	cgactacccgt	cgcacccccc	tcacccagga	4140
cggcgtctcg	cggctgcga	acgtcgcc	cgacacgtc	ttcatcattg	acgatgcgg	4200
cgagatcccg	gcccgtgacg	gcgtccctt	cctaagctc	tacgatggct	ggaccgagcc	4260
cttcgccttc	cacatgcgc	agtcgctcta	cggcttcttc	tggaaagcgc	cgggcaccc	4320
ggaggtgggt	tcaggctgca	cggtggacat	gctgcaggca	gtgtatgggc	tggacggcat	4380
ccgcctgcgc	cggcccgagt	actacaccat	gcccaacttc	agacagtatg	agaaccgcac	4440
cggccacatc	ctgggtcgagt	ggtcgctggg	cagccccctg	cacttcggcg	gctggcactg	4500
ctcctgtgc	ttcacgcccc	aggggatctt	cttcaagctc	gtgtccggcc	agaatggcga	4560
tttcccacgc	ttgggtgtact	acgaggacaa	gcgggacctg	aactacatcc	gcggcctgat	4620
ccgcacccggg	ggctgtttcg	acggcacgc	cgaggagatc	cgcgcgtc	accccagcga	4680
gcacatgtat	gcgccccagt	acgtcgtaa	gaactacgc	cggttccact	acctgcttgg	4740
caacccctac	caggagccca	ggagcacgc	ggcggggcg	tggcgcac	ggggtcccga	4800
gggaaggccg	ccgcggccgg	gcaaactgga	cgaggcggaa	gtcgaacaaa	aactcatctc	4860
agaagaggat	ctgaatttgg	atccttaggtt	taaactgagg	gcactgaatg	cgcttgcgt	4920
gctgaattgt	ttgtgtatgtt	ggtggcgat	tttggttaaa	taagtaagca	tggctgtat	4980
tttatcatat	gatcgatctt	ttgggttttta	tttaacacat	tgtaaaatgt	gtatctatta	5040
ataactcaat	gtataagat	tggttccatt	tcgggtgc	tagatctgt	tatttgcact	5100
gtgatgtttt	gactccaaaa	accaaataatca	aaactcaata	aactcatatg	atatgtccac	5160
ctgtttcttg	aagagttcat	ctaccattt	agttggcatt	tatcagtgtt	gcagcggcgc	5220
tgtgtttttt	aacataacaa	ttgttcacgg	catatatcca	cggccggcct	agctagccac	5280
ggtggccaga	tccactaggg	gcaagcggc	gcttaattaa	attnaatgt	ttaaactagg	5340
aaatccaagc	ttgggctgca	ggtcaatccc	attgttttg	aagcagctc	acattgtatc	5400
ctttctcgag	gtcatttata	tgcttgagaa	gagagtccgg	atagtccaaa	ataaaacaaa	5460
gttaagatata	cctggtcaaa	agtggaaaca	tcagttaaaa	ggtggatataa	agttttat	5520
cggtaataaa	agggtggccca	aagtggaaatt	tactctttc	tactattata	aaaattgagg	5580
atgtttttgt	cggtactttt	atacgtcatt	tttggatggaa	tttggttttt	agtttattcg	5640
cttttggaaa	tgcatatctt	tatttggatc	gggttttaag	ttcggttgc	tttggtaata	5700
cagaggaggat	tgtataagaa	atatctttaa	aaaaacccat	atgctaattt	gacataattt	5760
ttgagaaaaaa	tatataattca	ggcgaatttct	cacaatgaac	aataataaga	ttaaataatgc	5820
tttccccctgt	tgcagcgc	gggttattttt	tctagtaaaa	ataaaagata	aacttagact	5880
caaaacattt	acaaaaacaa	cccctaaatgt	tccttaaagcc	caaagtgc	tccacgatcc	5940
atagcaagcc	cagcccaacc	caaccccaacc	caacccaccc	cagtccagcc	aactggacaa	6000
tagtctccac	accccccac	tatcaccctg	agttgtccgc	acgcacccgc	cgttcgcag	6060
caaaaaaaaaa	aaaagaaaaag	aaaaaaaaaa	aaaaaaaagg	acagcaggat	ggtccgggtc	6120
gtggggccgg	gaaacgcgcg	gaggatcg	agccagcgc	gaggccgc	ctccctccgc	6180
ttccaaagaa	acgcccccca	tcgcccactat	atacatacc	ccccctctcc	tccatcccc	6240
ccaaccctac	caccaccacc	accaccaccc	ccacccctc	ccccctcgct	gccggacgac	6300
gcctcccccc	tccccctccg	ccggccgcgc	gccggtaacc	accccgcccc	tccctcttt	6360
ctttctccgt	ttttttttt	cgttcgcgtc	tcgatcttt	gccttggtag	tttgggtgg	6420
cgagaggccgg	cttcgtgcgc	gcccagatcg	gtgcgcggga	ggggcgggat	ctcgccgc	6480
ggggctctcg	cggcgtggat	ccggccggga	tctcgcgggg	aatggggctc	tcggatgtat	6540
atctcgatc	cggccgttgg	ggggggatgt	atggggggat	ttttttttt	ttttttttt	6600
acaagatcg	ggagggggaa	aaaggggact	atgggttata	ttttttatata	tttctgtgc	6660
ttcgtcaggc	tttagatgtgc	tagatcttt	tttcttctt	ttgtgggtag	aatttgaatc	6720
cctcagcatt	gttcatcggt	agttttttt	ttcatgattt	gtgacaaatg	cagcctcg	6780
cggagctttt	ttgttaggtag	accatggctt	ctccggagag	gagaccat	gagattaggc	6840
cagctacagc	agctgtatgt	gccgcgggtt	gtgatatcg	taaccat	attgagacgt	6900
ctacagtgaa	ctttaggaca	gagccacaaa	caccacaaga	gtggattat	gatctagaga	6960
ggttgtcaaga	tagatccct	ttgttgggtt	ctgagggtt	gggttggat	gctggatatt	7020
cttacgtctgg	gcccggaa	gcttaggaac	cttacgttgc	gacagtggat	agactgttt	7080
acgtgtcaca	taggtatccaa	aggttggcc	taggtatcc	ttgtacaca	cattgttta	7140
agtctatgaa	ggcgcacgg	tttaagttct	ttgttgcgt	tataggcctt	ccaaacgcac	7200
catctgttag	gttgcgtat	gttttggat	acacagcc	gggtacattt	cgccgcac	7260
gataacaagca	ttgggtatgg	catgtatgtt	gtttttggca	aagggat	gagttgcac	7320
ctccctccaag	gccagtttag	ccagttaccc	agatctgagg	taccctgac	tcgtgcgc	7380
cgtgtgcgtg	tccgtcgat	tttctggcc	ggcggggc	gggcgcgc	tcagaagcgt	7440
tgcgttggcg	tgtgtgtgt	tctgggttgc	tttaattttt	ccaagttgt	ttcaaggtgg	7500
atcgctgtgt	caaggccgt	gtgtttttt	gaccacccgg	cactggc	gagttgtgt	7560
gcttggatgt	gtttttttt	gtatgggtt	tatttgc	ttggatgtgt	gtactactt	7620
ggtttgttga	attttatgt	gcagttgcgt	attgtat	agctgggctt	cctggacatt	7680
gttatgtatt	aataaatgt	ttgcttctt	ctaaagatct	ttaagtgc	aattcatatt	7740
tcctcctgca	gggtttaaac	ttggccgtgc	ctat	aaagattccc	aatagtagtc	7800

12/15

caaaaatttt	gtaacgaagg	gagcataata	gttacatgc	aaggaaaaact	gccattttt	7860
agaggggatg	cttgtttaag	aacaaaaaaat	atatcactt	cttttgtcc	aagtcatgc	7920
gtatTTTTT	aaaaatattt	gttccttcgt	atattcgag	cttcaatcac	tttatggtc	7980
tttgtattct	ggcttgctg	taaatcgtag	ctaacccttct	tccttagcaga	aattattaat	8040
acttgggata	ttttttaga	atcaagtaaa	ttacatatta	ccaccacatc	gagctgcttt	8100
taaattcata	ttacagccat	ataggcttga	ttcattttgc	aaaatttcca	ggatattgac	8160
aacgttaact	taataatatc	ttgaaatatt	aaagctatta	tgatttaggg	tgc当地atgga	8220
ccgagtttgt	tcgggttata	tcaaaatcaa	accaaaccaa	ctatatcggt	ttggattgg	8280
tcggTTTTG	cgggtttca	gcattttctg	gtttttttt	tgttagatga	atattatTTT	8340
atcttactt	tgtcaaattt	ttgataagta	aatatatgt	tttagaaaaa	ttatTTTTT	8400
ttacaaacat	atgatctt	aaaaatattt	tataaggagaa	ttttcttaat	aacacatgt	8460
atTTTTTAT	tttagtcgtt	tgactaattt	ttcgttgcgt	tacacttca	aagttAACCA	8520
aatttagtaa	ttaagtataa	aaatcaatat	gatacctaaa	taatgatatg	ttctatTTAA	8580
ttttaaattt	tcgaaatttC	acttcaaatt	cggaaaagat	atataagaat	tttgatagat	8640
tttgacatAT	gaatatggaa	gaacaaagag	attgacgc	tttagtaaca	tttgataaga	8700
aagtgtatgt	acaaccaatt	atTTAAAGT	aataaaaatg	gagcacttca	tattaacga	8760
aatattacat	gccagaagag	tcgcaaataat	ttctagat	tttttaaaga	aaattctata	8820
aaaagtctt	aaggcatata	tataaaaact	atatattt	atTTGGTT	ggTCGAATT	8880
tgTTTACTC	aataccaaac	ttaaatttagac	caaataataat	ttggTTTTA	atcgcggccc	8940
actagtcccc	ggtgttagctt	ggcgtatca	ttggctatgc	tgtttccgt	gtggaaattgt	9000
tatccgctca	caattccaca	caacatacga	gccggaagca	taaagtgtaa	agcctgggt	9060
gcctaATGAG	tgagctact	cacattaatt	gcgttgcgt	cactgcccgc	tttccagtgc	9120
gaaaaACCTG	cgtccagct	gcattaatga	atcggcaac	gcgcggggag	aggcgggtt	9180
cgtattggc	gctttccgc	tgcgcacgct	gcgcacgctg	cgcacgc	ctcgctcact	9240
gactcgctc	gctcggtcg	tcggctgcgg	cgagcggtat	cagctcact	aaaggcggta	9300
atacggTTT	ccacagaatc	agggataac	gcggaaaga	acatgtgagc	aaaaggccag	9360
caaaaggcCA	ggaacggtaa	aaaggcccg	ttgtggcgt	ttttccatag	gctccggcccc	9420
cctgacgac	atcacaaaaa	tcgacgctc	agtcagaggt	ggcggaaaccc	gacaggacta	9480
taaagatacc	aggcgTTTCC	ccctggaagc	tccctcgtc	gctctccgt	ttcgaccctg	9540
ccgTTTACCG	gatacctgtc	cgccTTTCTC	ccttccggaa	gcgtggcg	tttcatacg	9600
tcacgctgt	ggtatctcag	ttcgggtgt	gtcgttgcgt	ccaagctggg	ctgtgtgcac	9660
gaaccccccg	ttcagcccga	ccgctgcg	ttatccggta	actatgtct	tgagtccaaac	9720
ccggtaagac	acgacttac	gccactggca	gcagccactg	gtaacagat	tagcagagcg	9780
aggtatgtag	gcgggtgtac	agagttctt	aagtgggtgc	ctaactacgg	ctacactaga	9840
aggacagtat	ttggTATCTG	cgctctgt	aaggcaggta	ccttgcggaa	aagagtgg	9900
agctctgtat	ccggcaaaac	aaccaccgct	ggtacgggt	tttttttgt	ttgcaagcag	9960
cagattacgc	gcagaaaaaa	aggatctaa	gaagatcctt	tgatTTTT	tacgggggt	10020
gacgctca	ggaacgaaaa	ctcacgtt	gggattttgg	tcatgagatt	atcaaaaagg	10080
atcttccac	agatcctttt	aaattaaaaa	tgaagtttt	aatcaatcta	aagtatata	10140
gagtaaactt	ggtctgacag	ttaccaatgc	ttaatcgt	aggcacctat	ctcagcgtatc	10200
tgtctattt	gttcatccat	agttgcctg	ctcccccgt	tgtagataac	tacgatacgg	10260
gagggcttac	catctggccc	cagtgtc	atgtacgt	gagacccacg	ctcaccggct	10320
ccagatttat	cgcaataaa	ccagccagcc	gggaggccg	agcgcagaag	ttgtcctgca	10380
actttatccg	cctccatca	gtctattaaat	ttgtccggg	aagctagat	aagttagtgc	10440
ccagttataa	gtttgcgaa	cgttgtg	attgtacag	gcatgtgt	gtcacgctcg	10500
tcgtttggta	tggTTTcatt	cagctccgg	tcccaacgt	caaggcgagt	tacatgtatcc	10560
cccatgttgt	gcaaaaaaagc	ggttagctc	ttcggctctc	cgatgtgt	cagaagtaag	10620
ttggccgcag	tgttatctact	catggttatg	gcagcactgc	ataattctct	tactgtcatg	10680
ccatccgtaa	gatgtttt	tgtgactgt	gagactcaa	ccaagtatt	ctgagaatag	10740
tgtatgcgc	gaccggatgt	ctcttgcgg	gcgtcaatac	gggataatac	cgcGCCACAT	10800
agcagaacat	taaaaatgt	catcttgg	aaacgttctt	ccccggcgaaa	actctcaagg	10860
atcttaccgc	tgttgatgt	cagttcgt	taacccactc	gtgcacccaa	ctgatcttca	10920
gcatctttt	ctttcaccag	cgTTTCTGG	tgagaaaaaa	caggaaggca	aaatgcccga	10980
aaaaaggggaa	taagggcgac	acggaaatgt	tgaataacta	tactcttct	ttttcaatat	11040
tattgaagca	tttatcaggg	ttattgtctc	atgagcggat	acatatttga	atgtatttag	11100
aaaaataaaac	aaatagggg	tccgcgcaca	tttccccgaa	aagtgcaccc	tgacgtctaa	11160
gaaaccat	ttatcatgac	attaacat	aaaaataggc	gtatcacgag	gccctttcgt	11220
ctcgcgcgt	tcgggtatg	cggtgaaaac	ctctgacaca	tgcagctccc	ggagacggtc	11280
acagcttgc	tgtacggcga	tgccggggagc	agacaagccc	gtcaggggcgc	gtcagcgggt	11340
gttggcgggt	gtcggggctg	gcttaactat	gcggcatcg	agcagattgt	actgagagtg	11400
caccatatgc	ggtgtaaat	accgcacaga	tgcgttaagga	aaaataccg	catcaggcgc	11460
cattcgccat	tcaggtgcgt	caactgttgg	gaagggcgt	cggtgcgggc	ctttcgct	11520
ttacgcccagc	tggcggaaagg	gggatgtgt	gcaaggcgt	taagttgggt	aacccagg	11580
ttttcccagt	cacgacgtt	taaaacgacg	gccagtgaat	tacaccgggt	tgatcatggg	11640
ccg						11643

<210> 12
<211> 115
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 12
catgattacg ccaagcttagc ggccgcattc ccgggaagct aggccaccgt ggcccgccctg 60
caggggaaagc ttgcatgcct gcagatcccc ggggatcctc tagagtgcac ctgca 115

<210> 13
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 13
gggtacccccc ggggtcgac 19

<210> 14
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 14
taatgagctc gtttaaa 17

<210> 15
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 15
cggccggcct agctagccac ggtggccaga tccactagtt ctagagcggc cgctt 55

<210> 16
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 16
cctgcagatc cccggggatc ctctagagtc gacctgca 38

<210> 17
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 17
gggtacccccc ggggtcgac 19

<210> 18
<211> 138

14/15

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 18
tggccaccgc ttaattaagg cgcgccatgc ccgggcaagc ggccgcttaa ttaaatttaa 60
atgtttaaac taggaatcc aagcttggc tgcaggtcaa tcccattgct tttgaagcag
ctcaacatg atctcttt 120
138

<210> 19
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 19
ggtaccctga gctc 34

<210> 20
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 20
gaattcatat ttcctcctgc agggttaaa cttgccgtgg c 41

<210> 21
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 21
cggccccacta gtcacccgtg t 21

<210> 22
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 22
gcgcacgctg cgcacgctgc gcacgct 27

<210> 23
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 23
acaccggtgt gatcatgggc cg 22

<210> 24
<211> 69
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 24
tggccaccgc ttaattaagg cgcgcacatgc cccctgcaga^tccccgggga tcctctagag
tcgacctgc 60
69

<210> 25
<211> 144
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 25
cggccggc^t agcttagccac ggtggccaga tccactaggg gcaagcggcc gcttaattaa
attnaaatgt ttaaactagg aaatccaagc ttgggctgca ggtcaatccc attgctttg
aagcagctca acattgatct ct^ttt 60
120
144

<210> 26
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 26
ggtagccctga gctc 14

<210> 27
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 27
gaattcatat ttcctcctgc agggttaaa cttgccgtgg c 41